



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA DOPRAVNÍ

Klára Egemová

Pokrytí hlavního města Prahy
systemem automatického počítání cestujících

Bakalářská práce

2024



K620 Ústav dopravní telematiky

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

Klára Egemová

Studijní program (obor/specializace) studenta:

bakalářský – DOS – Dopravní systémy a technika

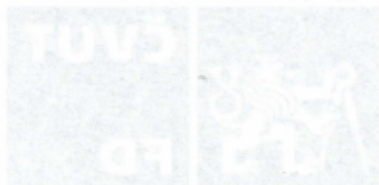
Název tématu (česky): **Pokrytí hlavního města Prahy systémem
automatického počítání cestujících**

Název tématu (anglicky): Automated Passenger Counting Coverage in Prague

Zásady pro vypracování

Při zpracování bakalářské práce se řiďte následujícími pokyny:

- analyzujte specifické vlastnosti APC dat
- analyzujte a vizualizujte stávající pokrytí Prahy vozidly vybavenými APC
- analyzujte chování parametru obsazenosti linky v čase
- formulujte doporučení pro práci s APC daty



Rozsah grafických prací: dle požadavků vedoucího práce

Rozsah průvodní zprávy: minimálně 35 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)

Seznam odborné literatury: VDV457 Automatic Passenger Counting Systems (APCS)
SQLite, <https://sqlite.org/index.html>
Vybraná zdrojová data: <https://pid.cz/o-systemu/.opendata/>

Vedoucí bakalářské práce:

Ing. Milan Sliacky, Ph.D.

Ing. Lukáš Hrdina

Datum zadání bakalářské práce:

27. září 2023

(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

Datum odevzdání bakalářské práce:

5. srpna 2024

- a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
- b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia

Ing. Zuzana Bělinová, Ph.D.
vedoucí
Ústavu dopravní telematiky



prof. Ing. Ondřej Příbyl, Ph.D.
děkan fakulty

Potvrzuji převzetí zadání bakalářské práce.

Klára Egemová
jméno a podpis studenta

V Praze dne 27. septembra 2023

Poděkování

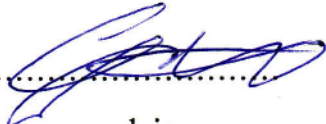
Na tomto místě bych ráda poděkovala Ing. Lukášovi Hrdinovi z Oddělení technického rozvoje a projektů ROPID p. o. za poskytnutý materiál, odborný dohled a velmi lidský přístup při konzultování postupu v rámci tvorby práce, děkuji. Zároveň děkuji Ing. Milanovi Sliackému, Ph.D. z Katedry dopravní telematiky Fakulty dopravní ČVUT za vedení práce a cenné připomínky.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracovala samostatně a uvedla jsem veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací a Rámcovými pravidly používání umělé inteligence na ČVUT pro studijní a pedagogické účely v Bc. a NM studiu.

Nemám závažný důvod proti užívání tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze dne 1. 8. 2024


.....
podpis

Abstrakt

Cílem této bakalářské práce je analýza stávajícího pokrytí hlavního města Prahy systémem automatického počítání cestujících (APC). V rámci analýzy je pracováno jak s daty získanými zařízeními APC ve vozidlech, tak s veřejně dostupnými jízdními řády GTFS. Dílčími závěry obsaženými v práci je vizualizace a zhodnocení pokrytí autobusového vozového parku PID, pokrytí území hlavního města Prahy, pokrytí oběhů v rámci času, pokrytí jednotlivých autobusových linek a rozložení obsazenosti vybraných linek v čase. Výsledkem analýzy je komplexní pohled na rozložení měřících zařízení APC v provozu a formulace doporučení pro práci s APC daty do budoucna.

Klíčová slova: automatické počítání cestujících, procento pokrytí, obsazenost linky, jízdní řády, databáze

Abstract

The aim of this bachelor's thesis is to analyze the current coverage of automated passenger counting systems (APC) in the city of Prague. The analysis utilizes data obtained from APC devices located in vehicles as well as GTFS timetables available for public use. The thesis includes visualization and evaluation of bus vehicle coverage, city-wide coverage in Prague, coverage of bus orders through time, coverage of individual bus lines, and the analysis of occupancy of selected lines over time. The result of this thesis is a comprehensive summary of the distribution of APC measuring devices and the formulation of recommendations for handling APC data in the future.

Key words: automated passenger counting, coverage percentage, line occupancy, timetables, database

Obsah

Seznam použitých zkratk	8
Úvod	9
1. Systémy APC	9
1.1 Druhy APC systémů	9
1.2 Společné předpoklady zařízení	10
1.3 Příklady APC zařízení	10
1.4 Obecné předpoklady v rámci PID	11
1.5 Systémy APC v provozu PID	11
1.5.1 Princip fungování APC systémů v provozu PID	12
1.5.2 Zařízení použité pro získání dat k analýze v rámci této práce	13
2. Specifické vlastnosti APC dat poskytnutých společností ROPID	13
2.1 Zpracovatelnost dat získaných systémem APC	13
2.2 Vhodnost dat získaných systémem APC pro analýzu	13
2.3 Předpoklady pro vytvoření doporučení do budoucna	14
2.4 Soubory obsahující analyzovaná data	14
2.5 Ověření kvality dat	16
3. Analýza a vizualizace stávajícího pokrytí Prahy vozidly vybavenými APC	18
3.1 Příprava dat	18
3.2 Vizualizace v mapě	22
3.3 Vybavení vozového parku systémem APC	26
3.3.1 Typy vozidel nasazených ve sledovaném období	28
3.4 Pokrytí jednotlivých linek v průběhu sledovaného období	29
3.4.1 Porovnání míry pokrytí jednotlivých linek	29
3.4.2 Detailní analýza pokrytí pořadí vybraných linek	31
3.5 Pokrytí oběhů systémem APC během dne	38
4. Chování parametru obsazenosti linky v čase	41
4.1 Obecný přístup k analýze	41
4.2 Linka 119	43
4.2.1 Obsazenost linky 119 v čase	44
4.2.2 Počet nastupujících cestujících v zastávce Navigátorů v čase	45
4.3 Linka 133	46
4.3.1 Obsazenost linky 133 v čase	47
4.3.2 Počet nastupujících cestujících v zastávce Pernerova v čase	48
4.4 Linka 201	49

4.4.1 Obsazenost linky 201 v čase.....	50
4.4.2 Počet nastupujících cestujících v zastávce Rajská zahrada v čase	51
4.5 Linka 352.....	52
4.5.1 Obsazenost linky 352 v čase.....	53
4.5.2 Počet nastupujících cestujících v zastávce Amforová v čase	54
5. Doporučení pro práci s APC daty.....	55
5.1 Vliv pokrytí na měření.....	55
5.2 Doporučení při získávání APC dat	55
5.2.1 Pokrytí vozového parku.....	55
5.2.2 Pokrytí linek	55
5.2.3 Pokrytí území hlavního města Prahy	56
5.2.4 Pokrytí v rámci času	56
5.3 Doporučení pro práci se získanými APC daty.....	57
5.3.1 Vypravení vozů s APC (apc-2023-02-06).....	57
5.3.2 Vybrané informace o vypravených spojích (apc-vypraveni_dpp)	57
5.3.3 Seznam jednotlivých typů vozů ve vozovém parku PID (seznam_busu)	58
5.3.4 Jízdní řády v podobě GTFS	58
Závěr.....	59
Zdroje	61
Seznam obrázků.....	62
Seznam tabulek.....	63
Seznam dotazů v jazyce SQLite	64
Seznam skriptů v jazyce R	66
Seznam příloh	72

Seznam použitých zkratk

2D	dvoudimenzionální
3D	trojdimenzionální
AI	Artificial intelligence
APC	Automated Passenger Counting / Automatické počítání cestujících
CAN	Controller Area Network
CSV	Comma-separated value
ČVUT	České vysoké učení technické v Praze
DB	databáze
DPP	Dopravní podnik hl. m. Prahy
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GTFS	General Transit Feed Specification
IBIS	Integrated On-Board Information System
IR	infrared
LCD	liquid-crystal display
LTE	Long Term Evolution
N/A	No value Is Available
OCC	odbavovací zařízení společnosti Mikroelektronika
PC	personal computer
PCU	Passenger Counting Unit
PID	Pražská integrovaná doprava
ROPID	Regionální organizátor Pražské integrované dopravy
RTD	Real Time Data
SQLite	zjednodušený Structured Query Language
TOF	Time-of-flight
VDV	Verband Deutscher Verkehrsunternehmen

Úvod

Automatické počítání cestujících je čím dál více využívaným nástrojem pro zlepšení kvality veřejné hromadné dopravy. Data ze systému APC mohou být zpracována za účelem různých výstupů, od zaznamenávání absolutních počtů cestujících v jednotlivých dopravních prostředcích, přes zkoumání obsazenosti dopravních prostředků v souvislosti s denní, týdenní či roční dobou, po plánování dopravní obslužnosti daného území s ohledem na kapacitu jednotlivých druhů dopravy. Jednání odborníků dle výsledků těchto analýz posléze vede ke zkvalitnění služeb, které cestujícím veřejná hromadná doprava nabízí.

Oproti ručnímu sčítání cestujících je provoz systému APC časově nenáročný, při dlouhodobém provozu finančně výhodnější a efektivnější při následném vyhodnocování získaných dat. Ať se jedná o ruční vozový nebo profilový průzkum, vždy je třeba počítat s možnou chybou člověka vzniklou z důvodu například nepřehlednosti situace či únavy (lidský faktor). Ani zařízení APC nemají přesnost 100 %, obvykle se dlouhodobá přesnost garantovaná dodavatelem pohybuje mezi 95 až 98 % [1]. Na rozdíl od ručního sčítání je však možno systém APC provozovat nepřetržitě a bez narůstajících nákladů.

Cílem této práce je zhodnocení stávajícího pokrytí hlavního města Prahy systémem APC. Sledovaným obdobím byl únor–červen 2023, a to výhradně dopravně inženýrsky významné dny: úterý, středa a čtvrtek. Pro analýzu byla uvažována data ze systému APC z celkem 65 dnů. Práce je blíže zaměřena na analýzu specifických vlastností APC dat, analýzu stávajícího pokrytí hl. m. Prahy vozidly autobusů vybavenými APC, analýzu chování parametru obsazenosti linky v čase a formulaci doporučení pro práci s APC daty. Data pro analýzu byla poskytnuta společností ROPID.

1. Systémy APC

1.1 Druhy APC systémů

Existuje mnoho podob zařízení APC. Každý typ zařízení APC má jiné předpoklady, a to jak hardwarové, tak softwarové. Zároveň je možné každý typ zařízení APC zakomponovat a spojit s centrálním systémem vozidla různými způsoby. Výrobci dodávající systémy APC z důvodu zachování konkurenceschopnosti detailní způsob fungování svých zařízení nezveřejňují. Zařízení mohou pracovat na základě analýzy videozáznamu, na principu čidla jako optické závory nebo na základě vysílání a odrazu světelných paprsků (infrared metoda). Velmi často výrobci těchto jednotek dodávají i vlastní software pro následné vyhodnocení a zobrazení získaných dat.

IR metoda a kamerová čidla mají oproti optické závoře mnoho výhod. Počítání cestujících metodou optické závory je v Praze využíváno již pouze v metru. Čidla umí rozlišovat směr průchodu, ale jinak sofistikovaná nejsou. Naopak záznamy z čidel fungujících na principu infračervených paprsků nebo na základě analýzy obrazu nesou mnohem větší množství údajů. S tím je spojena větší složitost dat a náročnější proces vyhodnocování. Čidlo je například schopné rozlišit, kdy cestující skutečně vystoupí a kdy pouze vystoupí, aby uvolnil cestu ostatním vystupujícím, a znovu nastoupí. V takovém případě čidlo nesmí cestujícího započítat dvakrát.

Také je třeba správně stanovit parametry, aby zařízení rozeznalo například velikost zavazadla, vyhodnotilo, že se jedná o zavazadlo, a určilo jeho kategorii dle naměřených rozměrů nebo aby dle naměřené výšky cestujícího poznalo, zda se jedná o dospělé osobu nebo dítě. Ne vždy jsou ale takto podrobné informace pro dopravce důležité – informace o kategorii zavazadel a cestujících může být vítaná, nikoliv však nezbytná. Záleží, za jakým účelem dopravce data ze systému APC sbírá.

1.2 Společné předpoklady zařízení

Posílání naměřených dat probíhá buď přes palubní PC nebo zařízení posílá data na svůj server samostatným modemem. Pro správné uložení dat zařízení APC potřebuje znát několik klíčových informací.

Musí umět zjistit svou aktuální polohu. To pozná zařízení APC buď z palubního počítače (z ručního vyhlášení zastávky nebo z GPS polohy), a nebo má přímo napojení na externí GNSS (nejčastěji GPS a Galileo).

Dále je třeba, aby zařízení poznalo, v jaké je právě zastávce – musí mít informace z jízdního řádu, aby dokázalo přiřadit spočítané cestující do příslušné zastávky. Je tady komplikace, kdy vůz zastaví dvakrát ve stejné zastávce, například popojede ke sloupku a znovu otevře dveře. Zařízení APC musí u druhého otevření dveří poznat, že je stále ve stejné zastávce. Další specifickou situací je najetí do konečné zastávky. Po výstupu všech cestujících se po vozidle pohybuje ještě řidič. Pokud má na konečné například přestávku, může i několikrát projít dveřmi ven a dovnitř a tyto průchody nesmí být započítány. Zařízení musí poznat, že má po výstupu cestujících přestat počítat.

V neposlední řadě je důležité, aby zařízení poznalo, že jsou otevřené dveře a má tedy počítat. Jsou dva způsoby detekce otevřených dveří. První způsob detekuje na základě dat získaných ze sběrnice autobusu, která sleduje stav dveří například z důvodu bezpečnosti (signalizace pro řidiče na palubní desce). Čidla APC jsou často dodělána do vozu později a mohou být na sběrnici napojena. Druhý způsob je aktivní zjištění stavu dveří, a to měřením napětí na osvětlení dveřního prostoru (svítí-otevřeno či nesvítí-zavřeno).

1.3 Příklady APC zařízení

Pro příklad jednotka sčítacího systému s názvem UCP 02 (výrobce Herman systems, s.r.o.) pracuje na základě snímání a následného zpracování obrazu [2]. Snímaný signál ze stereo kamery je zpracován a převeden na počet cestujících přímo v jednotce s čidlem (a tudíž není v rozporu se zákonem pro ochranu osobních údajů). Jiný způsob fungování zařízení pro sčítání cestujících využívá jednotka společnosti Mikroelektronika s.r.o., která pracuje na základě infračerveného světelného paprsku [3]. Čidlo rozpozná, zda cestující nastupuje nebo vystupuje a vnitřní elektronika na základě těchto údajů vypočítává a shromažďuje údaje o obsazenosti vozidla.

Současnou špičkou v kamerových systémech je model SC9133-RTL společnosti VIVOTEK Inc. [4]. Jako ve většině technických odvětvích, i v oblasti zařízení APC je využívána umělá inteligence. Toto zařízení je vybaveno nejnovější technologií AI Deep Learning pro detekci a klasifikaci lidí [5]. Díky této technologii výrobce garantuje až 99% přesnost. Zařízení dokáže rozeznat jednotlivé osoby v davu a dokáže je spolehlivě odlišit od jiných objektů. Rovněž dokáže identifikovat pohyb, který nepochází od lidí, a tak zabránit falešnému počítání.

Požadavky na zařízení a obsah odesílaných dat v rámci vozového parku PID jsou formulovány v souladu s normou VDV 457 (verze 4/2018) [6]. Druh ani fyzikální princip fungování zařízení není striktně definován. Dopravce rovněž využívá řešení v souladu se Seznamem certifikovaných zařízení pro provoz v PID, který je pravidelně zveřejňován na webu organizátora.

1.4 Obecné předpoklady v rámci PID

V dokumentu Standardy kvality pro autobusy systému PID [7] v podkapitole 4.2.7.2 Zařízení pro sčítání cestujících je uvedeno následující: „Vozidlo musí být vybaveno certifikovaným zařízením pro automatické sčítání cestujících s definovanou přesností měření. Rozsah vybavených vozidel dopravce je stanoven návaznou přílohou Odbavovací a informační zařízení ve vozidlech PID. Podmínkou je rovnoměrné využití vozidel se zařízením pro automatické počítání cestujících v rámci všech linek dopravce. Platí ode dne vyhlášení.“

Příloha Odbavovací a informační zařízení ve vozidlech PID [8] předchozího dokumentu v kapitole 4.10 Zařízení pro automatické sčítání cestujících dále specifikuje, že dopravce je povinen na začátku každého soutěženého kontraktu vybavit systémem APC minimálně 1/3 (33 %) vozového parku a každé další nově zařízené vozidlo do soutěženého svazku během trvání kontraktu musí být též zařízením vybaveno. Zároveň rozložení zařízení APC do různých typů vozidel by mělo být rovnoměrné v rámci plnění podmínky rovnoměrného sběru dat.

1.5 Systémy APC v provozu PID

Dle seznamu certifikovaných zařízení pro provoz v PID [9] společnost ROPID aktuálně užívá pět různých řešení systémů APC (k 29.11.2023). Žádost o certifikaci se posílá za celý systém, ne za čidlo, protože součástí certifikace je i přiřazování dat k jízdnímu řádu, a to čidlo samo nedělá.

První dva typy, APC PCU-230 a Vivotek SC8133, jsou dodána brněnskou společností Abirail CZ s.r.o. Komunikace obou typů zařízení s palubním počítačem probíhá přes rozhraní IBIS. Rozhraní IBIS slouží pro komunikaci prvků informačního systému ve vozidle a je doporučeným způsobem komunikace pro oba typy zařízení. Konkrétně jsou zařízení používána v kombinaci s OCC3 – Mikroelektronika spol. s r.o. (IBIS). Oba typy řešení jsou schváleny jak ve stávajících vozech, tak pro nové vozy.

Systém společnosti Abirail je obecně postavený tak, že čidla jsou přes Ethernet připojena k centrální jednotce Abirail, která má vlastní LTE modem a GNSS přijímač. Tato jednotka údaje vyhodnocuje a přiřazuje k zastávkám. Pro přiřazení využívá primárně GNSS polohy, ale pomáhá si odposlechem sběrnice IBIS palubního počítače OCC a zjišťuje, kdy OCC vyhlásilo zastávku. Velká část přiřazování probíhá v backoffice.

V seznamu se nachází další tři řešení APC. Systém od dodavatele Telmax s.r.o. označení PCU 1 je schválen pro stávající i nové vozy, momentálně probíhají úpravy na straně dodavatele a třetí strany. PCU 1 je souhrnné označení celého systému, v principu se jedná o čidla Iris IRMA Matrix přímo napojená na palubní počítač Telmax FCS2000, který provádí přiřazení k zastávkám a vyhodnocení.

Zbývající dva typy zapojení zařízení mají zažádáno o certifikaci a jsou testována přímo v provozu. Společnost Buse s.r.o. využívá čidla Iris IRMA Matrix připojená přes Ethernet k LCD panelu BUSE BS370. Přijímá zprávy o vyhlášení zastávky, lince, cíli, pořadí atd. z IBIS sběrnice palubního počítače OCC a data vyhodnocuje. LCD panel má vlastní modem (využívaný i pro přestupy), který využívá i pro přenos APC dat do backoffice.

APC systém společnosti KonekTel, a.s. označení OIS_APC1 rovněž využívá čidla IRMA Matrix. Ta jsou napřímo napojena do palubního PC výrobce ARBOR. Obě žádosti o certifikaci jsou ve fázi čekání na výzvu žadatele.

1.5.1 Princip fungování APC systémů v provozu PID

Všechny aktuálně používané typy APC zařízení v provozu PID fungují na stejném principu. Základem jsou senzory umístěné ve dveřích dopravního prostředku. U autobusů městské hromadné dopravy probíhá výstup a nástup cestujících všemi dveřmi, proto senzory musí být také umístěny ve všech dveřích. Nachází se ve stropě v rámu dveří vedle výstražného světla.



Obrázek 1 - Umístění zařízení IRMA Matrix [10]

Data nasbíraná senzory putují do vyhodnocovací jednotky a mohou pak být interpretována. Společnost Abirail CZ s.r.o. udává u svých řešení přesnost počítání nastupujících a vystupujících cestujících a zjišťování dalších doplňkových ukazatelů v nepřetržitém režimu s přesností 98% [1].

1.5.2 Zařízení použité pro získání dat k analýze v rámci této práce

Zařízením, z kterého jsou získána data pro analýzu, je model IRMA Matrix německé společnosti iris-GmbH infrared & intelligent sensor [10]. Zařízení lze instalovat a pomocí konektoru připojit k Ethernetu nebo CAN (Controller Area Network) a nakonfigurovat. Tento model je spolehlivý ve sběru dat pro následnou analýzu, zároveň je schopný poskytovat data i v reálném čase. Zatím není součástí žádného certifikovaného řešení.

Principem, na kterém model IRMA Matrix pracuje, je využití TOF (Time-of-flight) metody. Je to metoda založená na spočítání doby odrazu světelných paprsků (IR). Zařízení vysílá infračervené světlo a objekty v blízkém okolí ho odrážejí zpět k senzoru. Zařízení měří dobu, kterou odraz jednotlivých paprsků trvá („dobu letu“). Je pak schopno určit, jak daleko a kde se objekt nachází a skládá si 3D představu o svém okolí. Tato technologie je velice přesná a dokáže detekovat jak cestující, tak jejich zavazadla nebo další objekty nacházející se v zorném poli. Její výhodou je, že vlivy prostředí, jako jsou změny osvětlení, teplotní výkyvy nebo vzdušná vlhkost, nemají na šíření světla vliv, a tudíž nemohou mít negativní vliv na přesnost měření zařízení [11].

Nevýhodou je neschopnost detekovat osoby v blízkosti dveří. Zařízení nerozliší, kdy cestující skutečně vystoupí a kdy pouze vystoupí, aby uvolnil cestu ostatním vystupujícím a znovu nastoupí. To může mít za výsledek jednostranně ovlivněné výstupy měření.

2. Specifické vlastnosti APC dat poskytnutých společnostmi ROPID

2.1 Zpracovatelnost dat získaných systémem APC

Na problematiku dat ze systému APC může být pohlíženo z dvou odlišných úhlů pohledu. Jedna věc je, jak kvalitní a spolehlivé je pokrytí oběhů a druhá věc je, jak jsou kvalitní samotná získaná data. Jinými slovy, když bude mít zařízení APC 100% spolehlivost, tzn. nebude docházet ke zkreslení při získávání dat žádným způsobem, ale bude jimi vybaveno 1 % vozů z vozového parku, nelze očekávat, že získaná data budou jakkoliv směrodatná. Zajdu-li do druhého extrému, nemohu považovat za kvalitní výstup data zařízení APC, jimiž bude sice vybaveno 100 % vozů vozového parku, ale samotná přesnost zařízení bude značně nízká.

V rámci této práce jsou data získaná systémem APC analyzována dvěma různými přístupy. Nejprve je zdokumentováno pokrytí sledovaného území vozidly vybavenými zařízením APC. Bylo třeba zjistit, jakým způsobem a podle jakého klíče jsou vozidla nasazována, jestli je jich nasazován dostatek pro formulování závěrů nebo zda některé linky jsou pravidelně pokryty více než jiné. Poté následuje samotné vyhodnocování počtu nástupu a výstupu cestujících a informace s tím spojené, jako místo nástupu/výstupu, čas nástupu/výstupu a další.

2.2 Vhodnost dat získaných systémem APC pro analýzu

Vyhodnocování dat ze systému APC obecně probíhá na stejném principu. Je pracováno s daty z minulosti – jedná se o data, která jsou nejprve nějaký čas sbírána a ukládána, nikoli real-time data (RTD). Tato data jsou zanalyzována, obsahují velké množství detailů vztahujících se ke způsobu jejich pořízení, tudíž jsou vhodná pro vytvoření komplexních závěrů. Výstupem analýzy je zhodnocení stávajícího stavu a formulace předpokladů a doporučení do budoucnosti.

Pro zpracování jsou vhodná taková data, která neobsahují zbytečně mnoho úprav provozu linek a co možná nejméně dalších plánovaných změn v jízdním řádu. Data mají být reprezentativní co se týká běžného provozu a běžnému provozu by se měla co nejvíc blížit. Pro tuto práci by byla velmi nevhodná data například ze dnů jako jsou pátek a neděle nebo z období letních prázdnin. Tato práce popisuje metodu zpracování dat, která je demonstrována právě na ukázkových datech. Podstatou ale je, že tato metoda bude platná i pro širší záběr dat a bude možné ji aplikovat na jiné skupiny dat. Výběr dat je omezen z důvodu zjednodušení pro bakalářskou práci.

2.3 Předpoklady pro vytvoření doporučení do budoucna

Pro prezentování závěrů a formulaci doporučení do budoucna je třeba analýzu dat provést kvalitně a držet se zásad práce s velkým množstvím dat. Nejprve je třeba definovat problém, který má práce za cíl zkoumat. Následně je třeba získat data týkající se daného problému, taková data, jejichž analýzou je možno přiblížit řešený problém a podložit jej evidencí.

Předtím, než začne zpracování dat za účelem formulace závěru, je třeba ověřit a zajistit kvalitu dat. Nelze pracovat se souborem, který obsahuje chybné údaje. Ať už je nestandardní záznam označen v souboru dat, například v podobě vlajek v příslušných řádcích nebo hodnotami N/A (no value is available), nebo určité záznamy logicky nedávají smysl a chyba může být připsána stroji, tato data musí být odstraněna nebo jinak zajištěna jejich spolehlivost.

V moment, kdy jsou k dispozici relevantní data očištěná od chyb, lze začít jejich samotné zpracování za účelem smysluplného výstupu. Po provedení všech předešlých po sobě jdoucích kroků může přijít závěr, a to vyhodnocení skutečností, které byly během analýzy odhaleny a zadokumentovány.

2.4 Soubory obsahující analyzovaná data

Jako nejbližší období vhodné k analýze bylo určeno jaro 2023. Konkrétně se v práci věnuji datům získaným v období únor až červen roku 2023. Pro co možná nejvyšší spolehlivost dat s ohledem na jejich vypovídající hodnotu uvažuji pouze dopravně shodně se projevující dny: úterý, středa a čtvrtek. Ve výsledku se jedná o 5 měsíců, to je 22 týdnů, pro analýzu jsou uvažována data z 65 dnů.

Data zpracovaná v bakalářské práci vychází ze dvou hlavních zdrojů. Prvním zdrojem jsou data o vypravení vozů vybavených systémem APC poskytnuta společností ROPID v podobě CSV souboru (apc-2023-02-06). K tomuto souboru mi byly poskytnuty další tři soubory. První obsahuje vybrané informace o všech vypravených spojích pro jednotlivé dny (apc-vypraveni_dpp), druhý soubor obsahuje informace o jednotlivých typech vozů ve vozovém parku PID (seznam_busu) a třetí obsahuje vysvětlivky k jednotlivým hodnotám ve sloupci vlajky (význam_vlajek). Velikost souborů je uvedena v bajtech.

 apc-2023-02-06.csv	2,761,205,398
 apc-vypraveni_dpp.csv	2,111,779
 seznam_busu.csv	217,916
 význam_vlajek.txt	313

Obrázek 2 - Seznam souborů poskytnutých společností ROPID

Druhým hlavním zdrojem jsou volně dostupné jízdní řády v podobě GTFS pro dané období zveřejňované společností ROPID [12].

Soubor s názvem apc-2023-02-06 obsahuje 22 sloupců s údaji o jednotlivých spojih, například datum vypravení, evidenční číslo vozidla, čas začátku a konce směny, linku, uzel a zastávku definující jednoznačné místo nahlášení, počet cestujících ve vozidle a další. Každý řádek se pak liší místem ohlášení záznamu. Ohlášení záznamu probíhá vždy v jednotlivých zastávkách po výstupu a nástupu cestujících po konečném zavření dveří vozidla (po ukončení zastávky – poté, co vozidlo opustí určený prostor zastávky). Důležitým sloupcem je sloupec vlajky, který značí, zda je záznam v pořádku, nebo byl vygenerován s chybou. Z tohoto souboru může být výstupem například počet APC spojů skrze jednotlivé zastávky nebo celkový počet vypravených spojů s APC za konkrétní den.

Soubor obsahující počty všech vypravených spojů pro jednotlivé dny obsahuje pouze sloupce linka, pořadí, vůz a datum. Tento soubor je potřebný pro získání poměru pokrytých a nepokrytých oběhů systémem APC a poměru vozidel se systémem APC a bez. Není ale možné jej využít pro detailnější analýzu, protože neobsahuje údaje o poloze záznamů ani jiné další upřesňující informace.

Jízdní řády ve formátu GTFS jsou vydávány na dva týdny dopředu a primárně nejsou vázány na jednotlivá data. Každý soubor GTFS je složka obsahující 16 textových souborů. Výsledkem propojení těchto souborů na základě společných identifikačních označení, které obsahuje každá tabulka, je jednoznačný jízdní řád pro veřejnou hromadnou dopravu v Praze a Středočeském kraji. Velikost souborů je uvedena v bajtech.

agency.txt	162
calendar.txt	35,634
calendar_dates.txt	281,729
fare_attributes.txt	509
fare_rules.txt	471
feed_info.txt	164
levels.txt	4,006
pathways.txt	53,617
route_stops.txt	626,722
route_sub_agencies.txt	44,567
routes.txt	85,125
shapes.txt	142,541,799
stop_times.txt	101,871,245
stops.txt	1,187,887
transfers.txt	357,324
trips.txt	5,658,120

Obrázek 3 - Ukázka obsahu složky GTFS s platností na 2 týdny

Přestože jízdní řády GTFS mají platnost 14 dní, je třeba je vydávat každý den. Organizace veřejné hromadné dopravy musí reagovat na stále se měnící a vyvíjející dopravní situaci, a tak často dochází ke změnám jízdních řádů. Jedná se například o plánované výluky, neplánované výluky, reakce na operativní odklony, zavádění nových opatření a další. Z tohoto důvodu byl postup následující: sledované období únor–červen 2023 se skládá z 22 týdnů, z každého týdne nás pak zajímají 3 dny. Každá trojice dnů je navázána na jeden GTFS soubor vydaný v datum prvního z těchto tří dnů. Pro příklad měsíc únor:

Tabulka 1 - Ukázka výběru souborů GTFS pro jednotlivá data (datum)

GTFS 01.02.2023	GTFS 07.02.2023	GTFS 14.02.2023	GTFS 21.02.2023	GTFS 28.02.2023
	07.02. - úterý	14.02. - úterý	21.02. - úterý	28. 02. - úterý
01.02. - středa	08.02. - středa	15.02. - středa	22.02. - středa	01.03. - středa
02.02. - čtvrtek	09.02. - čtvrtek	16.02. - čtvrtek	23.02. - čtvrtek	02.03. - čtvrtek

Volně dostupná je též dokumentace k jízdním řádům GTFS. Obecné informace o textových souborech a konkrétní specifikace GTFS jsou dostupné na oficiálních stránkách Google transit [13], podrobné informace o konkrétním významu každého údaje na stránkách Pražské integrované dopravy [14].

2.5 Ověření kvality dat

Ověření kvality dat je provedeno u souboru s názvem apc-2023-02-06. Kontrola je dvojí. První jsou odstraněny záznamy, které mají nevyhovující hodnotu ve sloupci Vlajky – hodnota je jiná než 1. Takový údaj značí, že záznam byl upraven nebo s ním bylo jinak manipulováno. V souboru se kromě hodnoty 1 vyskytlo dalších 8 možností. Jelikož zpětně nelze přesně určit, z jakého důvodu byl záznam upraven, je pro spolehlivost analýzy na místě tyto záznamy vůbec neuvažovat. Problém by nastal, kdyby procento chybových dat přesáhlo 5 % celku. Dle společnosti ROPID je akceptovatelná chyba právě do výše 5 %. Toto procento se vztahuje k celkové chybě napočítaných cestujících a chybně uložený záznam může být jedním z důvodů. Pokud by chybovost dat překročila 5 %, bylo by nutné zvolit jinou strategii při kontrole kvality dat, například se blíže zabývat zkoumáním jednotlivých významů vlajek a důvody, proč se u záznamu objevily.

Z celkových 21 695 326 záznamů byla hodnota vlajky 1 u 20 965 077 záznamů a vlajka jiná než hodnoty 1 byla u 730 249 záznamů (viz Dotaz 1), to tvoří 3,37 %. Závěrem je, že 96,63 % záznamů je v pořádku.

Tabulka 2 - Spolehlivost všech obdržených záznamů

APC data netříděná	
Celkový počet zastavení	21 695 326
Vlajka v záznamu je 1	20 965 077
Vlajka v záznamu není 1	730 249
Spolehlivost dat	96,63 %

Bližší analýza výskytu vlajek je provedena na užším výběru dat. Z původního souboru jsou dotazem v jazyku SQLite vybrány všechny záznamy, které byly pořízeny v den úterý, středa a čtvrtek (viz Dotaz 2). Poté jsou uloženy do nového souboru a je s nimi dále pracováno. Tento soubor je dále nazýván Vypravení vozů s APC. Z nového celkového počtu 10 432 349 záznamů je vlajka v záznamu rovna 1 u 10 081 905 záznamů a vlajka jiné hodnoty než 1 u 350 444 záznamů. Tato vlajková data tvoří 3,36 % celku, zdravá data tvoří 96,64 % z celkového počtu záznamů. Nejčtenější hodnotou vlajky jinou než 1 byla vlajka 129 (178 597 záznamů – 1,71 %) – poškozený záznam, chyba snímače. Spolehlivost dat je téměř totožná, jako u předchozí analýzy. Z toho plyne, že vlajkové záznamy jsou napříč daty rozděleny rovnoměrně a nedochází například k vyšší chybovosti ve více vytižených dnech.

Tabulka 3 - Spolehlivost vybraných dnů, rozpis počtu záznamů s hodnotou vlajky jinou než 1

APC data výběr út + st + čt		
Celkový počet zastavení	10 432 349	
Vlajka v záznamu je 1	10 081 905	
Vlajka v záznamu není 1	350 444	
vlajka 5	18 992	generovaný záznam
vlajka 9	3 324	zastávka upravena
vlajka 18	49 762	časy upraveny
vlajka 34	48 610	počty cestujících upraveny
vlajka 58	42 193	celý záznam upraven (nebo sloučen z více záznamů)
vlajka 62	8 473	záznam upraven – rozdělení záznamu konečné zastávky
vlajka 129	178 597	poškozený záznam – chyba snímače
vlajka 133	493	poškozený záznam
Spolehlivost dat	96,64 %	

Druhou kontrolou je odstranění dat, která logicky nedávají smysl. Ke zkreslení dat dojde zpravidla chybou měřicího přístroje. Zde bylo uvaženo odstranění záznamů, kde počet nastupujících nebo vystupujících cestujících byl naměřen příliš vysoký nebo nižší než 0 (záporný počet cestujících nelze naměřit). Žádná taková data se v souboru nenacházela. Rovněž je kontrolován počet cestujících, kteří se ve vozidle nacházeli v jeden okamžik. Nejkapacitnější používaný autobus v daném období je SOR NB 18 City s délkou 18 metrů a maximální kapacitou 152 cestujících [15]. Maximální naměřená obsazenost autobusu je 156 osob, což po zvážení může být reálná hodnota.

U složek GTFS kontrola dat není v rámci této práce proveditelná. Důvodem je to, že data ve složkách nejsou výstupem ze zařízení, tudíž neobsahují sloupec vlajky. Zároveň data není možné s ničím porovnat a říct, co dává nebo nedává smysl, takže jsou brána jako spolehlivá. Během analýzy se vyskytl problém s dvěma linkami, 873 a 877, které v GTFS jízdním řádu nebyly uvedeny, ačkoliv byly uvedeny v souboru Vypravení vozů s APC. V obou případech jde o mimořádné linky, nebyly v provozu po celé zkoumané období a jejich vyloučením ze souboru Vypravení vozů s APC se zvýší celková přesnost závěrů analýzy.

3. Analýza a vizualizace stávajícího pokrytí Prahy vozidly vybavenými APC

3.1 Příprava dat

Jedním z cílů této práce je vizualizace stávajícího pokrytí města Praha vozidly vybavenými systémem APC. Pracováno je se souborem Vypravení vozů s APC a s příslušnými soubory GTFS. Ze souboru Vypravení vozů s APC bylo třeba získat počet spojů vybavených systémem APC přes jednotlivé zastávky, ze souboru GTFS je pak získán celkový počet spojů přes každou zastávku. Následným porovnáním těchto dvou hodnot pro každou zastávku je vypočítáno procentuální pokrytí dané zastávky. Pro každou zastávku je možno zobrazit průměrný počet spojů za den, z toho průměrný počet spojů za den vybavených systémem APC a příslušné procento pokrytí zastávky systémem APC.

Soubor Vypravení vozů s APC je zpracován v databázi SQLite skrze program DB Browser [16]. Dotazem je pro každý den získána tabulka s počty spojů, které byly uskutečněny přes každou zastávku a byly vybaveny systémem APC (viz Dotaz 3).

```
1 SELECT distinct nazev, zast, uzel, COUNT(*) AS count
2 FROM apc_final_ut_st_ct
3 WHERE datum = '2023-02-01'
4 GROUP BY uzel || '_' || zast
5 ORDER BY nazev, zast;
```

Obrázek 4 - Dotaz 3, dotaz pro počet APC spojů, které byly uskutečněny přes každou zastávku

Dotaz vybírá jedinečné kombinace sloupců nazev, zast a uzel, potom spočítá, kolikrát se daná kombinace v souboru vyskytuje pro určené datum a tuto hodnotu uloží do sloupce s názvem count (počet). Dotaz se musí pro každé datum manuálně upravit a spustit (řádek 3). Výsledná tabulka je pro 01. 02. 2023 uložena pod názvem „APC_02_01_vystup“ a stejným způsobem pro každé další datum.

Dále je třeba získat celkový počet spojů (tripů) přes každou zastávku. Z každé GTFS složky je pro tuto úlohu potřeba 6 textových souborů, jimi jsou: calendar, calendar_dates, routes, stop_times, stops a trips. Pro soubor calendar je klíčový údaj service_id. Pro každý service_id je přiřazen údaj, které dny je vypraven a začátek a konec platnosti jízdního řádu.

Soubor calendar_dates je rovněž navázán na service_id, pro vybrané service_id udává plánovanou změnu v pravidelném jízdním řádu a to datem, v kterém bude změna platná a exception_type 1 nebo 2, 1 pokud je spoj nad rámec přidán a 2, pokud spoj mimořádně nepojede.

Soubor routes je navázán přes údaj route_id (číslo linky) a pro dotaz je důležitý sloupec route_type. Zde jsou filtrovány záznamy, kde route_type pro daný route_id je rovno 3, to znamená, že se jedná o autobus.

Soubor stop_times je první uvedenou tabulkou v dotazu, na kterou jsou navázány ostatní přes sloupec stop_id.

Soubor stops nese údaje o jednotlivých uzlech a zastávkách, například geografickou polohu. Důležité je pohlídat, aby u vybíraných záznamů byla hodnota location_type 0 – aby šlo o standardní zastávku, nikoli například speciálně označený vstup/výstup z metra.

Posledním souborem je soubor trips, který je navázán přes sloupec trip_id a obsahuje též sloupce service_id a route_id. Z důvodu dostupnosti dat je zde nutno pohlídat, aby byly vybrány pouze záznamy, které mají ve sloupci sub_agency_id číslo 1 – jedná se o Dopravní podnik hl. m. Prahy.

Dotaz je komplexní a hlídá veškeré mimořádné změny v provozu zanesené do složky GTFS (viz Dotaz 4).

```

1 SELECT
2     stops.stop_name,
3     stops.asw_node_id,
4     stops.asw_stop_id,
5     COUNT(DISTINCT trips.trip_id) AS trip_count
6 FROM
7     stop_times
8 JOIN
9     trips ON stop_times.trip_id = trips.trip_id
10 JOIN
11     stops ON stop_times.stop_id = stops.stop_id
12 JOIN
13     calendar ON trips.service_id = calendar.service_id
14 JOIN
15     routes ON trips.route_id = routes.route_id
16 LEFT JOIN
17     calendar_dates ON (calendar_dates.service_id = calendar.service_id AND calendar_dates.date = '20230201')
18 WHERE
19     calendar.start_date <= '20230201'
20     AND calendar.end_date >= '20230201'
21     AND ((calendar.wednesday = true AND (calendar_dates.date IS NULL OR (calendar_dates.date != '20230201' AND calendar_dates.exception_type = '2'))))
22     OR (calendar_dates.date = '20230201' AND calendar_dates.exception_type = '1'))
23     AND routes.route_type = 3
24     AND trips.sub_agency_id = 1
25     AND stops.location_type = 0
26 GROUP BY
27     stops.stop_name,
28     stops.asw_node_id,
29     stops.asw_stop_id;

```

Obrázek 5 - Dotaz 4, dotaz pro získání počtu všech spojů přes každou zastávku

Výstupem dotazu je seznam zastávek, u každé její název, jednoznačné označení polohy (uzel a směr) a počet všech spojů, které zastávkou projely v konkrétní datum. Jelikož GTFS složky je nutné stahovat postupně a názvy jednotlivých textových souborů v nich se na základě data neliší, je pro každou trojici dnů zvlášť vytvořena databáze. Při spuštění dotazu je vždy nutno změnit datum na pěti místech (řádky 17, 19, 20, 21 a 22) a příslušný den v týdnu (řádek 21). Tento seznam je pro 01.02.2023 uložen pod názvem GTFS_01_02_vystup a stejným způsobem pro každé další datum.

Následuje krok, kde je pro každou zastávku porovnán počet všech spojů a počet spojů vozidel se systémem APC. Dotaz pracuje se soubory získanými v předchozích krocích a opět je nutno spouštět pro každé datum zvlášť (viz Dotaz 5).

```

1 SELECT
2     stop_name,
3     GTFS_02_01_vystup.asw_node_id,
4     GTFS_02_01_vystup.asw_stop_id,
5     CAST(trip_count AS INTEGER) AS trip_count_numeric,
6     CAST(APC_02_01_vystup.count AS INTEGER) AS apc_count_numeric,
7     COALESCE(ROUND(((CAST(APC_02_01_vystup.count AS INTEGER) /
8     NULLIF(CAST(GTFS_02_01_vystup.trip_count AS FLOAT), 0)) * 100, 2), 0) AS apc_percentage
9 FROM
10     GTFS_02_01_vystup
11 LEFT JOIN
12     APC_02_01_vystup ON GTFS_02_01_vystup.asw_node_id = APC_02_01_vystup.uzel
13                     AND GTFS_02_01_vystup.asw_stop_id = APC_02_01_vystup.zast;

```

Obrázek 6 - Dotaz 5, dotaz pro porovnání pokrytých a nepokrytých spojů

Výsledná tabulka je pro 01.02.2023 uložena pod názvem POKRYTI_02_01 a stejným způsobem pro každé další datum.

	stop_name	asw_node_id	asw_stop_id	trip_count_numeric	apc_count_numeric	apc_percentage
	Filter	Filter	Filter	Filter	Filter	Filter
1	Ametystová	1274	1	31	NULL	0
2	Amforová	1029	1	255	193	75.69
3	Amforová	1029	2	251	189	75.3
4	Anděl	1040	1	18	1	5.56
5	Anděl	1040	2	15	1	6.67
6	Anděl	1040	4	8	NULL	0
7	Anděl	1040	11	5	NULL	0
8	Anděl	1040	12	297	102	34.34
9	Anděl	1040	14	295	100	33.9
10	Antala Staška	950	1	339	192	56.64
11	Antala Staška	950	2	151	101	66.89
12	Apolinářská	956	1	62	NULL	0
13	Apolinářská	956	2	62	NULL	0
14	Arbesovo náměstí	4	3	110	31	28.18
15	Arbesovo náměstí	4	4	110	33	30
16	Archeologická	356	1	191	2	1.05
17	Archeologická	356	2	191	NULL	0
18	Argentincká	48	1	10	NULL	0

Obrázek 7 - Ukázka souboru POKRYTI_02_01

V tento moment máme 65 souborů s názvem POKRYTI_XX_XX pro každý den a zbývá spočítat průměrné pokrytí každé zastávky systémem APC za celé zkoumané období a výsledek vizualizovat v mapě. Tento úkol je zpracován v programu RStudio [17] v jazyku R [18]. Jednak proto, že pro zpracování velkého množství souborů je přehlednější než databázové programy a jednak proto, že v něm existují různé knihovny, které jsou velmi vhodné pro zobrazování dat v mapě.

Prvně je provedeno vypočítání průměrného pokrytí zastávek za 65 dnů. Jednotlivé tabulky POKRYTI_XX_XX jsou formou inner join propojeny mezi sebou, ve výsledném pokrytí jsou zahrnuty pouze ty zastávky, které jsou společné pro všech 65 dnů, aby výsledek měl vypovídající hodnotu (viz Skript 1).

```

1 library(dplyr)
2 library(tidyr)
3
4 stops <- read.csv("stops.csv")
5 POKRYTI_02_01 <- read.csv("POKRYTI_02_01.csv")
6 POKRYTI_02_02 <- read.csv("POKRYTI_02_02.csv")
7 ...
8 POKRYTI_06_29 <- read.csv("POKRYTI_06_29.csv")
9
10 result_df <- POKRYTI_02_01 %>%
11   inner_join(POKRYTI_02_02, by = c("stop_name", "asw_node_id", "asw_stop_id")) %>%
12   inner_join(POKRYTI_02_07, by = c("stop_name", "asw_node_id", "asw_stop_id")) %>%
13   ...
14   inner_join(POKRYTI_06_29, by = c("stop_name", "asw_node_id", "asw_stop_id"))
15
16 result_df <- result_df %>%
17   inner_join(stops, by = c("stop_name", "asw_node_id", "asw_stop_id")) %>%
18   mutate(apc_vysledny = rowMeans(select(., apc_percentage), na.rm = TRUE),
19          tripu_zaden_prumer = rowMeans(select(., trip_count_numeric), na.rm = TRUE),
20          apc_zaden_prumer = rowMeans(select(., apc_count_numeric), na.rm = TRUE)) %>%
21   replace_na(list(apc_zaden_prumer = 0)) %>%
22   select(stop_name, asw_node_id, asw_stop_id, stop_lat, stop_lon, apc_vysledny, tripu_zaden_prumer, apc_zaden_prumer) %>%
23   distinct() %>%
24   arrange(stop_name, asw_node_id, asw_stop_id)
25
26 write.csv(result_df, "Procento_pokryti.csv", row.names = FALSE)

```

Obrázek 8 - Skript 1, skript pro vypočtení procenta pokrytí pro jednotlivé zastávky

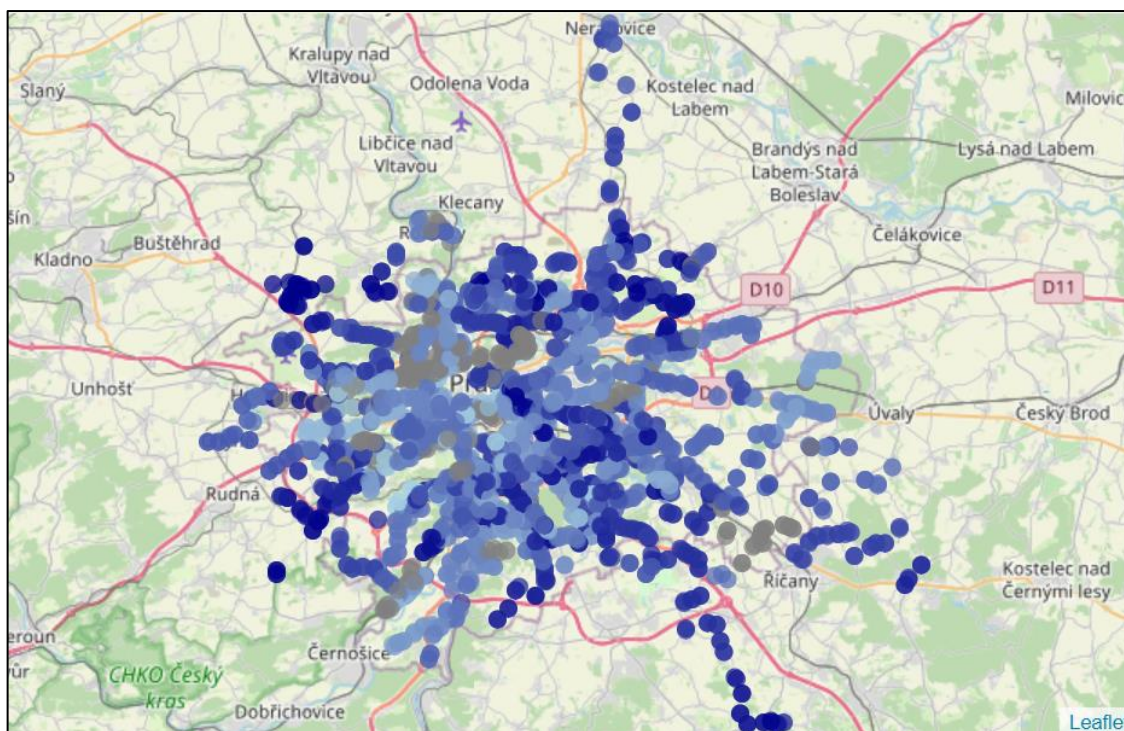
Výsledek je uložen do souboru Procento pokrytí, který obsahuje veškeré údaje potřebné pro vizualizaci dat. Název zastávky, její číselné označení a polohu, příslušné procento pokrytí systémem APC zaokrouhlené na dvě desetinná místa, průměrný počet všech spojů za den a průměrný počet spojů se systémem APC za den.

	stop_name	asw_node_id	asw_stop_id	stop_lat	stop_lon	apc_vysledny	tripu_za_den_prumer	apc_za_den_prumer
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	Ametystová	1274	1	49.988201	14.362217	0	31	0
2	Amforová	1029	1	50.041851	14.327284	69.81	265	185
3	Amforová	1029	2	50.041706	14.327312	71.05	266	189
4	Anděl	1040	1	50.07193	14.403631	5.56	18	1
5	Anděl	1040	2	50.071953	14.402808	13.33	15	2
6	Anděl	1040	4	50.070976	14.404552	25	8	2
7	Anděl	1040	11	50.071735	14.401952	20	5	1
8	Anděl	1040	12	50.071499	14.402929	34.86	327	114
9	Anděl	1040	14	50.071518	14.40265	35.91	323	116
10	Antala Staška	950	1	50.041695	14.444311	63.52	392	249
11	Antala Staška	950	2	50.041519	14.444415	64.06	384	246
12	Apolinářská	956	1	50.07151	14.42648	0	62	0
13	Apolinářská	956	2	50.07151	14.427031	0	62	0
14	Arbesovo náměstí	4	3	50.076405	14.405704	45.05	111	50
15	Arbesovo náměstí	4	4	50.076561	14.405448	45.95	111	51
16	Archeologická	356	1	50.043568	14.336884	11.16	215	24
17	Archeologická	356	2	50.044666	14.338751	9.77	215	21
18	Argentinská	48	1	50.103676	14.443435	0	19	0
19	Argentinská	48	2	50.102432	14.443085	0	47	0

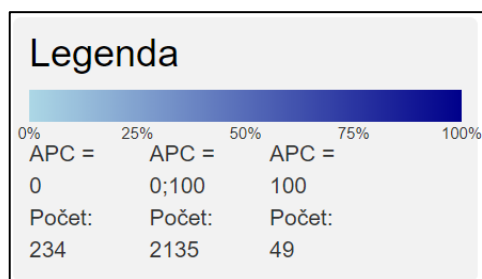
Obrázek 9 - Ukázka souboru Procento pokrytí

3.2 Vizualizace v mapě

Soubor Procento pokrytí je zdrojem údajů pro vizualizaci dat v mapě. Interaktivní verzi mapy možno spustit v programu RStudio (viz Skript 2).

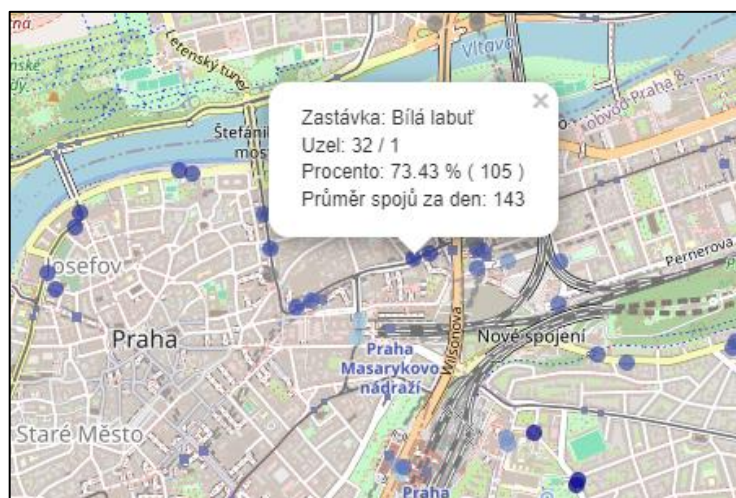


Obrázek 10 - Vizualizace pokrytí jednotlivých zastávek s legendou, statický výřez interaktivní mapy



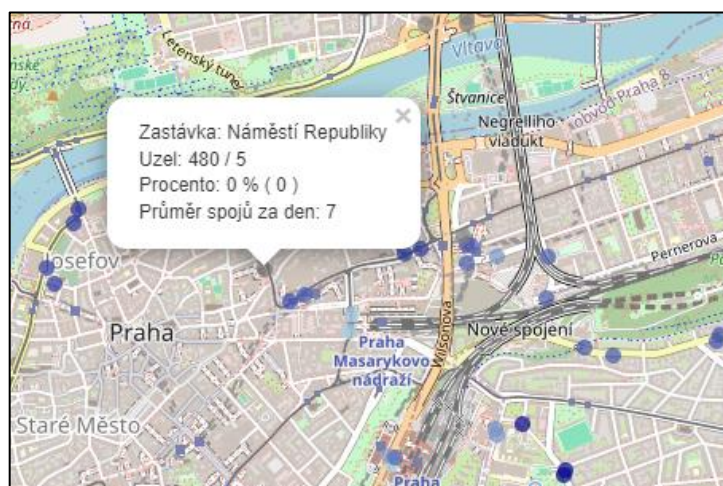
Výsledné pokrytí je spočítáno a zobrazeno pro 2 418 zastávek. Z toho 234 z nich neprojel za sledované období ani jeden spoj vybaven systémem APC a pokrytí těchto zastávek je tedy 0 % (šedá barva). Naopak 49 ze sledovaných zastávek má procento pokrytí 100 %, tedy všechny spoje, které zde projely, systém APC měly. Zbýlých 2 135 zastávek se procentem pokrytí nachází v intervalu mezi předchozími krajními hodnotami.

Po přiblížení je možno ke každé zastávce zobrazit konkrétní hodnoty, které jsou reprezentovány sytostí barvy. Například zastávkou Bílá Labuť směr Náměstí Republiky projelo ve sledovaném období průměrně 143 autobusových spojů za den, z toho průměrně 105 spojů bylo vybaveno systémem APC, a tak výsledné pokrytí zastávky je 73,43 %.



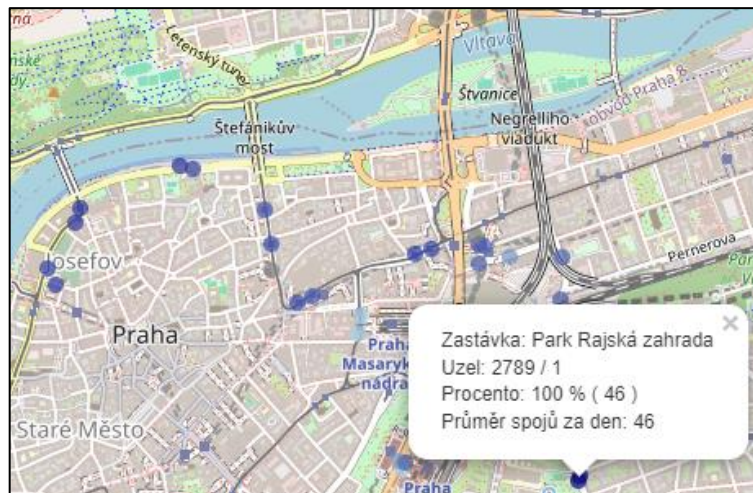
Obrázek 11 - Výřez mapy pokrytí, zastávka Bílá labuť, uzel 32/1

Naopak uzel 480/5 zastávky Náměstí republiky má výsledné pokrytí 0 %. Průměr spojů za den je roven 7 a navíc zastávka je pouze v jednom směru. Z těchto informací by se dalo usoudit, že se nejedná o pravidelnou zastávku, možná jde o jinak nestandardní zastávku, například noční na znamení. Ve skutečnosti se jedná o zastávku pro speciální linku H1 (289) pro osoby s omezenou schopností pohybu a orientace. Důvodem nulového pokrytí je to, že pro linku jsou vyhrazena starší vozidla bez systému APC.



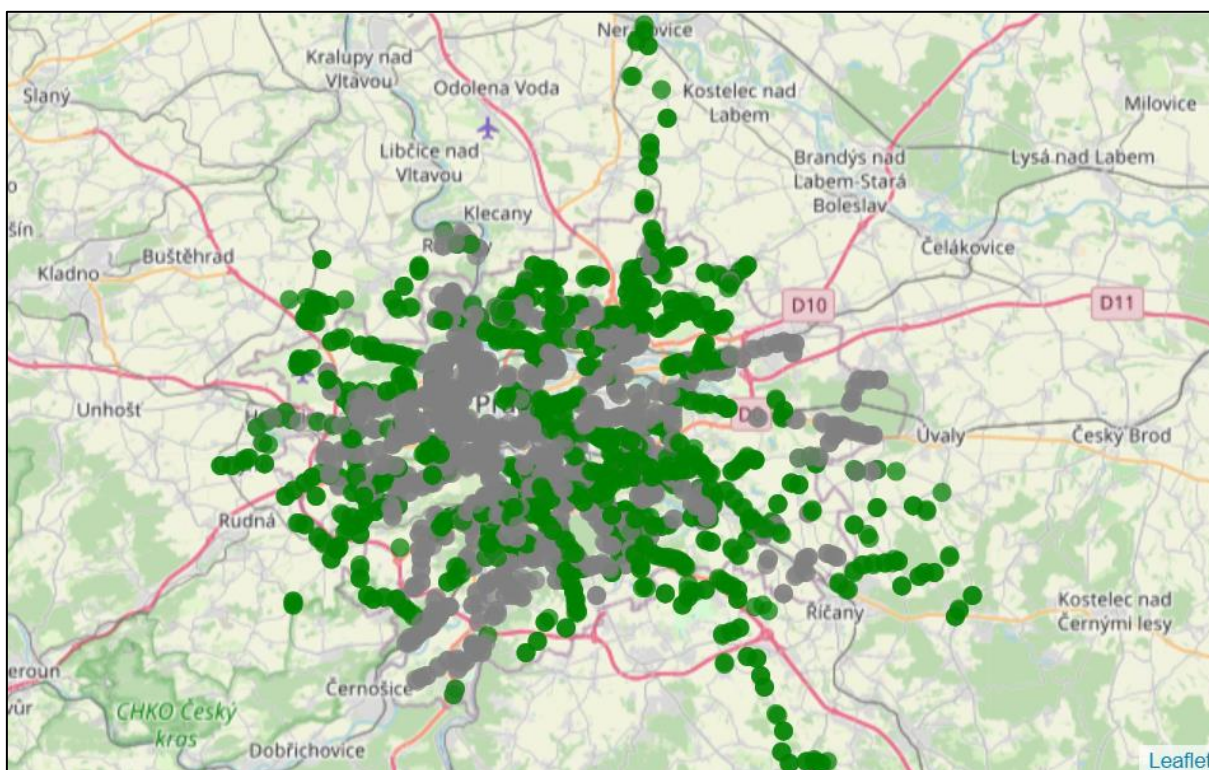
Obrázek 12 - Výřez mapy pokrytí, zastávka Náměstí Republiky, uzel 480/5

Zastávka Park Rajská zahrada ve směru Náměstí Winstona Churchilla má procento pokrytí 100 %, z celkových průměrných 46 autobusových spojů za den byly systémem APC vybaveny všechny. Bližší analýzou je zjištěno, že zastávkou projíždí pouze linka 101, na kterou jsou pravidelně nasazována vozidla typu Urbino 10.5 IV. Tato vozidla byla do vozového parku nově dodána (2020-2021) [19] a všechna disponují systémem APC.



Obrázek 13 - Výřez mapy pokrytí, zastávka Park Rajská zahrada, 2789/1

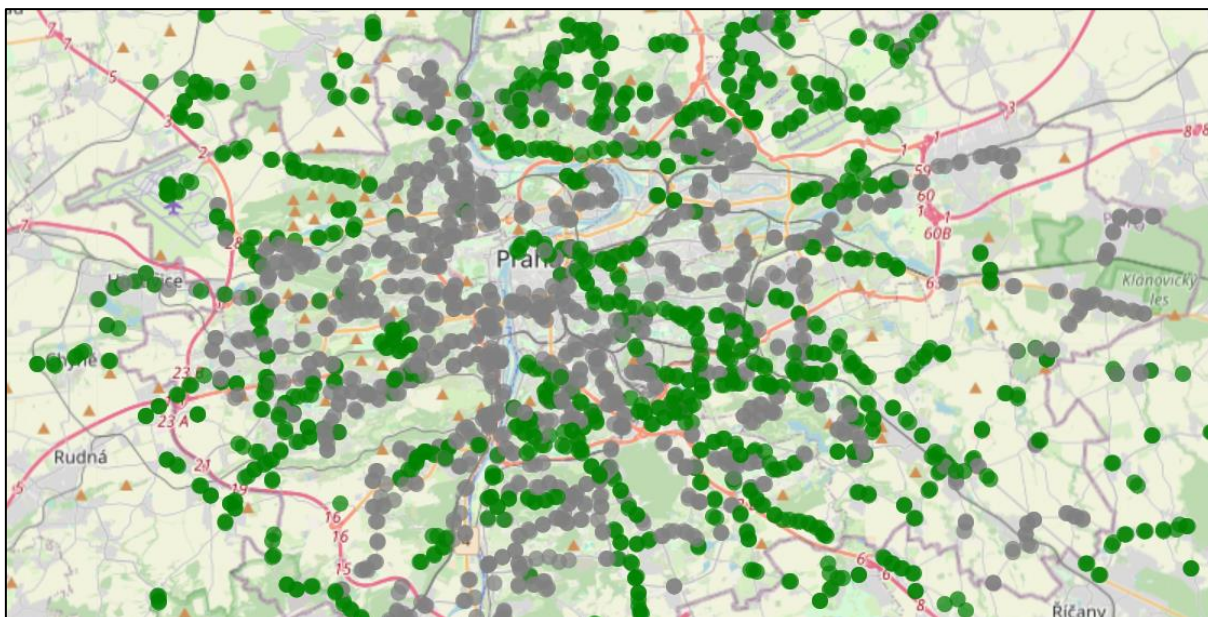
Ve stejné mapě můžeme též zobrazit rozdělení zastávek podle toho, zda procento pokrytí systémem APC dosáhlo 50 procent či ne (viz Skript 3). Právě 50 procent pokrytí je pro odborníky společnosti ROPID hranice pro vypovídající hodnotu výstupu ze systému APC a je pak tedy možné dále zpracovávat údaje, které systém APC zaznamenává.



Obrázek 14 - Vizualizace zastávek dle výše pokrytí (hranice 50 %) s legendou, statický výřez interaktivní mapy

Legenda	
Procento < 50:	Počet:
	1251
Procento ≥ 50:	Počet:
	1167

Z celkového počtu 2 418 zastávek je procento pokrytí menší než 50 u 1 251 z nich, procentem vyšším než 50 je pokryto 1 167 z nich. Tedy přibližně polovina zastávek je vhodná pro bližší analýzu dat ze systému APC.



Obrázek 15 - Výřez mapy pokrytí, centrum

Přiblížení na širší centrum Prahy ukazuje, že je centrum Prahy procentuálně méně pokryté než okraje města.

3.3 Vybavení vozového parku systémem APC

Nejprve je dotazem ze souboru Vypravení vypsán seznam dnů, které se v souboru vyskytují, jako kontrola, že není počítáno s nějakými dny navíc (viz Dotaz 6). Dat je 65 a shodují se se sledovanými dny (data myšleno ve smyslu slova datum). Ze souboru Vypravení vozů s APC je vygenerován seznam jednotlivých pořadí vozidel se systémem APC seřazený podle data, linky a pořadí (viz Dotaz 7).

```
1 select distinct datum, linkmen, poradi from apc_final_ut_st_ct
2 order by datum asc, linkmen asc, poradi asc;
```

Obrázek 16 - Dotaz 7, dotaz pro získání pořadí

Seznam je uložen pod názvem „linkmen_poradi_s_apc“ a pomocí funkce left join je propojen s tabulkou vypravení následujícím způsobem (viz Dotaz 8).

```
1 SELECT v_datum, v_linka, v_poradi, km_linkmen, km_poradi
2 FROM vypraveni
3 LEFT JOIN linkmen_poradi_s_apc
4 ON vypraveni.v_datum = linkmen_poradi_s_apc.km_datum AND
5     vypraveni.v_linka = linkmen_poradi_s_apc.km_linkmen AND
6     vypraveni.v_poradi = linkmen_poradi_s_apc.km_poradi
7 WHERE v_linka > 99;
```

Obrázek 17 - Dotaz 8, dotaz pro propojení souborů pořadí

Nyní změnou řádku 7 získáme informaci o počtu pokrytých a nepokrytých oběhů. Pokud poslední podmínka je ponechána prázdná, je získán celkový počet vypravených spojů za zkoumané období. Tento počet je 62 504.

```

1 SELECT v_datum, v_linka, v_poradi, km_linkmen, km_poradi
2 FROM vypraveni
3 LEFT JOIN linkmen_poradi_s_apc
4 ON vypraveni.v_datum = linkmen_poradi_s_apc.km_datum AND
   vypraveni.v_linka = linkmen_poradi_s_apc.km_linkmen AND
   vypraveni.v_poradi = linkmen_poradi_s_apc.km_poradi
7 WHERE v_linka > 99 --AND km_linkmen is null;

```

Obrázek 18 – Dotaz 8, dotaz pro získání seznamu vypravených pořadí

Pokud jsou spočítány řádky, které se nacházejí pouze v tabulce vypraveni, je získáno číslo 30 937, což je počet oběhů, které nebyly vybaveny systémem APC. Naopak pokud necháme spočítat pouze řádky, které se nachází jak ve vypravení, tak v tabulce „linkmen_poradi_s_apc“, získáme číslo 31 567, a to je počet oběhů, které byly systémem APC vybavené.

Tabulka 4 - Počty pokrytých a nepokrytých pořadí

Vypravení	
Celkový počet vypravených pořadí	62 504
Bez APC	30 937
S APC	31 567
Procento pokrytí pořadí	50,50 %

Závěrem je, že více než polovina vypravených oběhů systémem APC vybavena byla. Záměrně říkám oběhů, a ne vozidel, jelikož je možné, že některá vozidla byla za den vypravena vícekrát, a tedy použita znovu. Pokud by vozy vybavené systémem APC byly záměrně používány více než vozy bez systému APC, může být procento pokrytí oběhů vyšší, než by bylo procento vybavenosti vozového parku vozy se systémem APC. Tuto domněnku je možno potvrdit následující sérií dotazů.

Do databáze je importován soubor seznam_busu obsahující seznam autobusů, ze kterého nás zajímají autobusy ve vozovém parku Dopravního podniku hl. m. Prahy. Dotazem je získán seznam evidenčních čísel vozidel, která se alespoň jedenkrát objevila v souboru Vypravení (viz Dotaz 9). Na to je napojen soubor Vypravení vozů s APC přes evidenční číslo vozidla a spočítáno, kolik různých evidenčních čísel autobusu se v tabulce objeví, když je přidána nebo naopak smazána podmínka, že navázané evč je prázdné (řádek 5).

```

1 SELECT DISTINCT bus_actual, evč
2 FROM seznam_pouzitych_busu
3 LEFT JOIN apc_final_ut_st_ct
4 ON seznam_pouzitych_busu.bus_actual = apc_final_ut_st_ct.evč
5 WHERE apc_final_ut_st_ct.evč is not NULL
6 ORDER BY bus_actual ASC;

```

Obrázek 19 - Dotaz 9, dotaz pro získání seznamu použitých vozidel

Počet všech evidenčních čísel, která se vyskytla v datech, tudíž počet různých vozidel, která byla ve sledovaném období vypravena, je 1 188, z toho bez systému APC je 628 vozidel. Systémem APC bylo vybaveno 561 vozidel, což z celkového počtu tvoří 47,22 %.

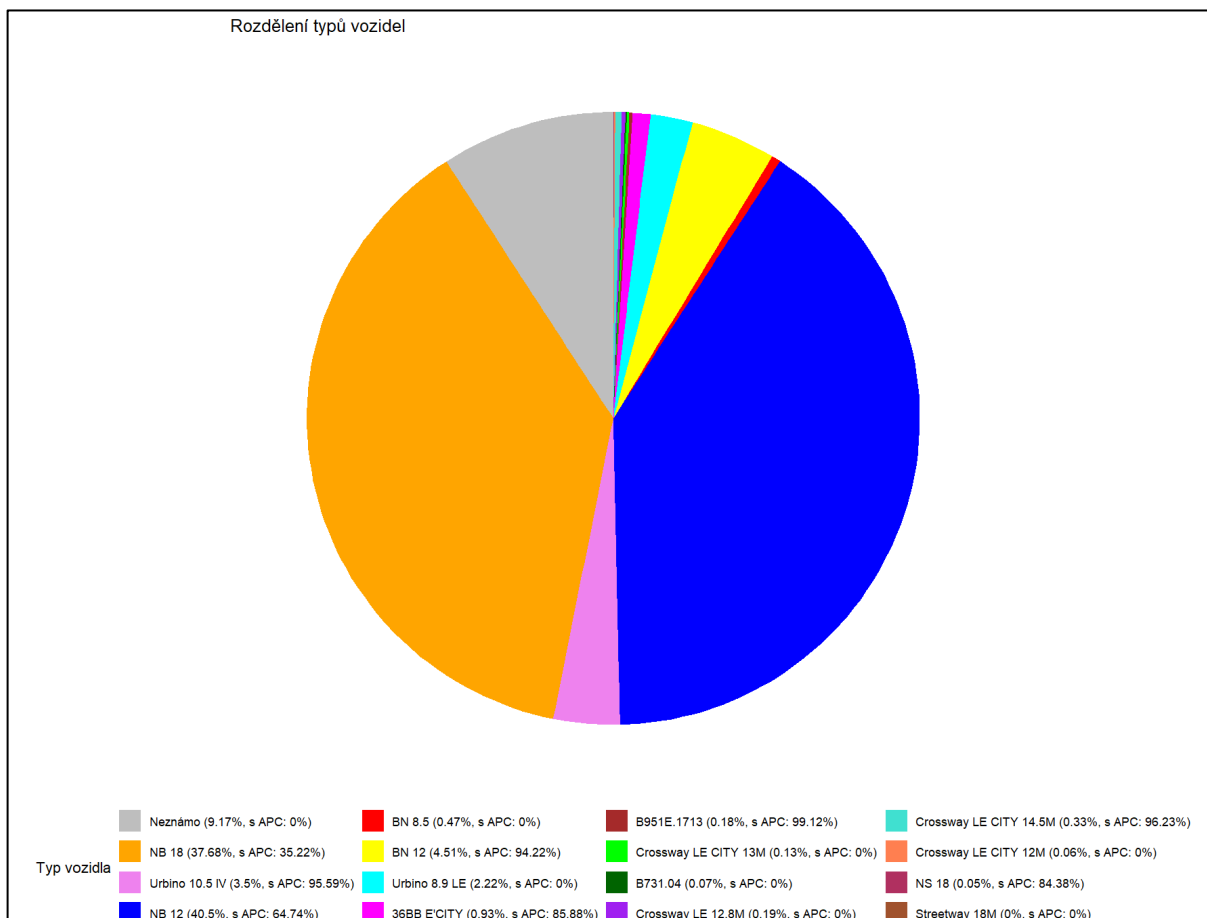
Tabulka 5 - Pokrytí použitých vozidel autobusů

Použitá vozidla autobusů	
Celkový počet vypravených vozidel autobusů	1188
Bez APC	628
S APC	561
Procento vybavení	47,22 %

Procento pokrytí užitého vozového parku je o 3,28 procentního bodu nižší než procento pokrytí vypravených oběhů. To znamená, že skutečně některé vozy vybavené systémem APC byly vypraveny vícekrát za den. Procento vybavení vozového parku 47,22 % o 14,22 procentního bodu převyšuje minimální požadavek na pokrytí vozového parku 33 % uvedený v návazné příloze Odbavovací a informační zařízení ve vozidlech PID.

3.3.1 Typy vozidel nasazených ve sledovaném období

Vypravená vozidla je možno rozdělit na 16 kategorií, co jedna kategorie, to jeden typ. Procentuální zastoupení typů v rámci celkového počtu 1 188 vozidel je zobrazeno v koláčovém grafu (viz Skript 4). V legendě je ke každému typu uvedeno procento zastoupení (první číslo v popisce) a procento vybavenosti daného typu systémem APC (druhé číslo v popisce).



Obrázek 20 - Graf rozdělení typů vozidel ve vozovém parku PID použitých v daném období

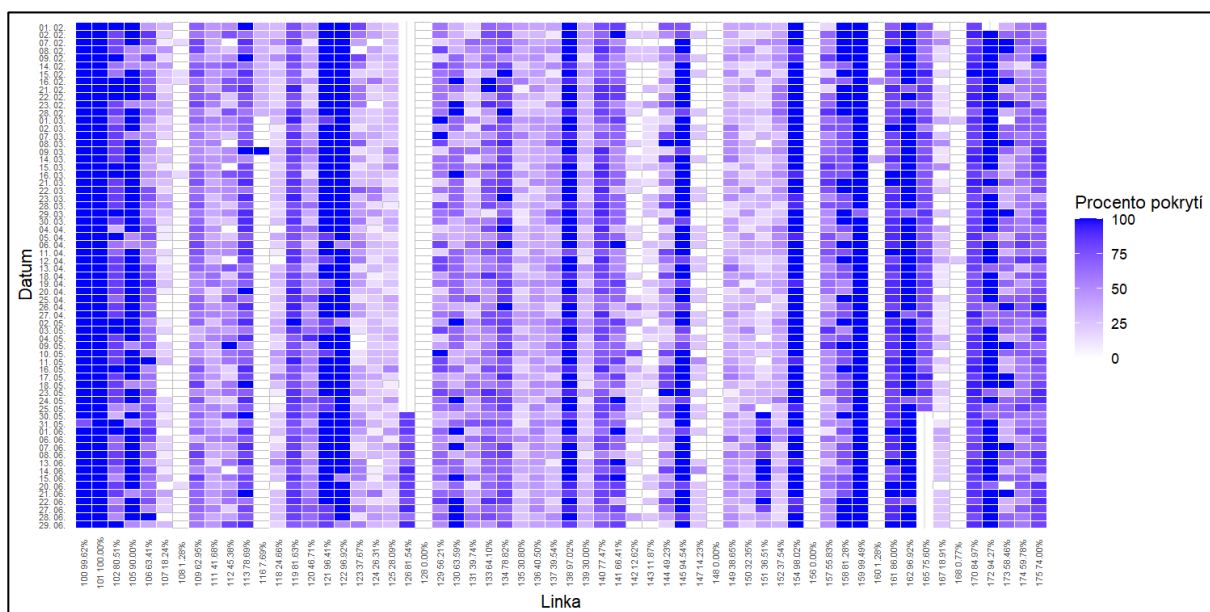
Nejzastoupenější jsou kategorie NB 18 (zastoupení 37,68 %, z toho vybaveno APC 35,22 %) a NB 12 (zastoupení 40,5 %, z toho vybaveno APC 64,74 %). Objevilo se zde několik záznamů o nasazení autobusů typu B951E.1713 a B731.04. Jde však o chybu v databázi. Jedná se o tři muzejní vozidla výrobce Karosa, kterým byla po vyřazení z běžného provozu ponechána jejich evidenční čísla, která se nyní shodují s evidenčními čísly nových vozů [20]. Ve skutečnosti tedy vozy Karosa daném období v provozu nebyly.

3.4 Pokrytí jednotlivých linek v průběhu sledovaného období

3.4.1 Porovnání míry pokrytí jednotlivých linek

Cílem vizualizace pokrytí jednotlivých linek v průběhu sledovaného období je zjistit, zda byly během období linky pokryty systémem APC rovnoměrně, nebo nerovnoměrně, popřípadě s pravidelným opakováním. Vizualizace je provedena v podobě 2D grafu. Na ose x jsou vyneseny všechny linky, na ose y jsou zobrazené jednotlivé dny, tj. celkem 65 dnů. Pro každou linku a každý den by bylo vypočítáno procentuální pokrytí jednotlivých pořadí systémem APC. Pokud všechna pořadí byla pokryta, tak výsledné pokrytí je 100 %, pokud nebylo pokryto žádné pořadí, tak 0 %. Tato hodnota byla převedena na barevnou škálu od bílé (0 %) po tmavě fialovou (100 %).

V programu DB Browser (databáze SQLite) byly provedeny dotazy pro získání seznamu všech vypravených oběhů (pořadí) a rovněž byla získána informace, zda vozidlo daného pořadí bylo vybaveno systémem APC nebo ne (viz Dotaz 8). Tento výstup byl uložen v podobě CSV souboru, znovu načten v programu RStudio a doplněn o výpočet procent pokrytí. Za pomoci knihoven dplyr a ggplot2 byl výstup převeden do grafické podoby (viz Skript 5).



Obrázek 21 - Pokrytí jednotlivých linek v průběhu sledovaného období s legendou, prvních 60 linek (celý graf viz Příloha 1)

U každé linky je též vypočítáno celkové průměrné pokrytí za celé sledované období. Toto procento je zobrazeno v popisu osy x pro každou linku. Linek se 100% pokrytím bylo v provozu celkem 5. Jsou to linky 101, 355, 872, 900 a 902. Z těchto linek ale pouze linky 101, 355 a 902 byly vypraveny každý ze sledovaných dnů. Linka 872 byla v provozu jen 5 dnů, linka 900 byla v provozu jen 3 dny.

Linek, které nebyly systémem APC pokryty vůbec nebo téměř vůbec, tedy výsledné procento je pod hodnotou 1, je 14. Z těchto linek byly v provozu každý den linky 128 (0 %), 148 (0 %), 156 (0 %), 168 (0,77 %), 216 (0,77 %), 245 (0,77 %) a 359 (0,77 %). Linka 292 (0 %) byla v provozu jeden den, linka 820 (0 %) byla v provozu 18 dnů, linka 891 (0 %) jeden den, linka 893 (0 %) jeden den, linka 894 (0 %) jeden den, linka 896 (0 %) jeden den a linka 898 (0 %) dva dny.

Linky začínající číslicí 8 nejsou uvedeny v souhrnu linek PID provozovaných v Praze nebo zajižďujících do Prahy [21]. Nejedná se o linky pravidelné, bývají to zpravidla linky zavedené dočasně, například kvůli výluce nebo jiné nestandardní dopravní události. Proto většina z nich nebyla v provozu po celé sledované období.

Linek, které měly celkové pokrytí nižší než 5 % a jedná se o linky pravidelné, bylo v provozu 13. K výše zmíněným 7 linkám s procentem pokrytí pod hodnotou 1 (linky 128, 148, 156, 168, 216, 245 a 359) je to dále linka 108 (1,28 %), linka 160 (1,28 %), linka 909 (1,03 %), linka 911 (3,46 %), linka 912 (1,54 %) a linka 916 (3,08 %).

Pravidelných linek s celkovým pokrytím vyšším než 95 % bylo v provozu 11. K výše zmíněným 3 linkám se 100% pokrytím (linky 101, 355 a 902) je to linka 100 (99,62 %), linka 121 (96,41 %), linka 122 (96,92 %), linka 138 (97,02 %), linka 154 (98,02 %), linka 159 (99,49 %), linka 162 (96,92 %) a linka 347 (96,92 %).

3.4.2 Detailní analýza pokrytí pořadí vybraných linek

Detailní analýza jednotlivých linek je provedena na základě CSV souboru nesoucím pro každou skupinu datum, linka a pořadí informaci o tom, zda toto pořadí bylo pokryto systémem APC či nikoliv, evidenční číslo vozidla na tomto pořadí a typ tohoto vozidla. Tento soubor byl získán dotazem v databázi spojením několika tabulek (viz Dotaz 10).

```
1 WITH bus_data AS (
2     SELECT
3         v_datum, v_linka, v_poradi, apc_linkmen, apc_poradi, v_vuz, apc_evč,
4         CASE
5             WHEN v_vuz IS NULL AND apc_evč IS NOT NULL THEN apc_evč
6             WHEN apc_evč IS NULL AND v_vuz IS NOT NULL THEN v_vuz
7             WHEN v_vuz IS NULL AND apc_evč IS NULL THEN NULL
8             WHEN v_vuz = apc_evč THEN apc_evč
9             ELSE 'CHYBA'
10        END AS bus_actual
11    FROM
12        vypraveni
13    LEFT JOIN
14        linkmen_s_apc_s_evc ON vypraveni.v_datum = linkmen_s_apc_s_evc.apc_datum
15                          AND vypraveni.v_linka = linkmen_s_apc_s_evc.apc_linkmen
16                          AND vypraveni.v_poradi = linkmen_s_apc_s_evc.apc_poradi
17    WHERE v_linka > 99
18 )
19 SELECT
20     bus_data.v_datum, bus_data.v_linka, bus_data.v_poradi, bus_data.apc_linkmen,
21     bus_data.apc_poradi, bus_data.v_vuz, bus_data.apc_evč, bus_data.bus_actual,
22     CASE
23         WHEN bus_actual = apc_evč THEN 'A'
24         WHEN bus_actual = 'CHYBA' THEN 'A_CHYBA'
25         ELSE 'N'
26     END AS melo_apc,
27     COALESCE (
28         CASE
29             WHEN bus_actual = 'CHYBA' THEN (
30                 SELECT seznam_busu.type
31                 FROM seznam_busu
32                 WHERE seznam_busu.number = bus_data.apc_evč
33             )
34             ELSE seznam_busu.type
35         END, 'Neznámo') AS type
36    FROM
37        bus_data
38    LEFT JOIN
39        seznam_busu ON bus_data.bus_actual = seznam_busu.number
40    WHERE
41        CASE
42            WHEN bus_actual = apc_evč THEN 'A'
43            WHEN bus_actual = 'CHYBA' THEN 'CHYBA'
44            ELSE 'N'
45        END IN ('A', 'N', 'CHYBA')
46    ORDER BY
47        v_datum, v_linka, v_poradi;
```

Obrázek 22 - Dotaz 10, dotaz pro získání tabulky obsahující informace o vypravených autobusech

Takto je vytvořena tabulka, která je dále využita jako zdroj dat pro vizualizaci pokrytí oběhů jednotlivých linek.

	v_datum	v_linka	v_poradi	apc_linkmen	apc_poradi	v_vuz	apc_evč	bus_actual	melo_apc	type
13	2023-02-01	102	2	102		NULL	3979	3979	A	NB 12
14	2023-02-01	102	3	102	3	3892	3892	3892	A	NB 12
15	2023-02-01	102	4	102	4	3989	3989	3989	A	NB 12
16	2023-02-01	102	5	102	5	NULL	3812	3812	A	NB 12
17	2023-02-01	105	1	105	1	2312	2312	2312	A	Urbino 10.5 IV
18	2023-02-01	105	2	105	2	NULL	3833	3833	A	NB 12
19	2023-02-01	106	1	106	1	NULL	4057	4057	A	NB 12
20	2023-02-01	106	2	NULL	NULL	3607	NULL	3607	N	NB 12
21	2023-02-01	106	3	NULL	NULL	3675	NULL	3675	N	NB 12
22	2023-02-01	106	4	NULL	NULL	NULL	NULL	NULL	N	Neznámo
23	2023-02-01	106	5	106	5	NULL	3917	3917	A	NB 12
24	2023-02-01	106	6	NULL	NULL	NULL	NULL	NULL	N	Neznámo
25	2023-02-01	106	7	106	7	NULL	4080	4080	A	NB 12
26	2023-02-01	107	1	107	1	NULL	5079	5079	A	NB 18
27	2023-02-01	107	2	NULL	NULL	6937	NULL	6937	N	NB 18
28	2023-02-01	107	3	107	3	5090	5090	5090	A	NB 18

Obrázek 23 - Ukázka tabulky obsahující informace o vypravených autobusech

V každém řádku je vyplněn sloupec v_datum, v_linka a v_poradi, pokud jsou vyplněny i sloupce apc_linkmen a apc_poradi, znamená to, že dané pořadí bylo pokryto, v opačném případě pořadí nebylo pokryto a oba sloupce mají hodnotu NULL.

Ve sloupci v_vuz se nachází evidenční číslo vozidla pro daný oběh získané z tabulky vypravení. Z pravděpodobně technických důvodů někdy záznam o evidenčním čísle chybí.

Ve sloupci apc_evč je evidenční číslo vozidla pro daný oběh získané ze souboru Vypravení vozů s APC. V ideálním případě by se toto evidenční číslo mělo shodovat s evidenčním číslem z předchozího sloupce. V takovém případě je toto evidenční číslo zobrazeno v následujícím sloupci bus_actual udávajícím, jaké vozidlo skutečně bylo nasazeno pro daný oběh. V případě, že se evidenční číslo nachází pouze v jednom ze sloupců v_vuz a apc_evč (důvodem jsou nejspíše technické problémy při přenosu dat), je jako bus_actual přepsáno toto číslo. Pokud se evidenční číslo nenachází ani v jednom ze sloupců, sloupec bus_actual zůstane prázdný (nemáme záznam).

Může nastat i situace, že oba sloupce obsahují evidenční číslo vozidla a tato čísla jsou rozdílná. To znamená, že pro dané pořadí bylo nasazeno jiné vozidlo, než jaké bylo plánované, a přišlo se na to až z důvodu, že se tato dvě evidenční čísla neshodují. V takovém případě se ve sloupci bus_actual objeví CHYBA.

Následuje sloupec melo_apc nesoucí informaci o tom, zda na daný oběh bylo nasazeno vozidlo se systémem APC (ve sloupci „A“) nebo bez něj (ve sloupci „N“). Pokud ve sloupci bus_actual je CHYBA, bude ve sloupci melo_apc „A_CHYBA“, protože evidentně existuje záznam o tomto oběhu v souboru Vypravení vozů s APC a tak vozidlo muselo být systémem APC vybaveno, přestože podle vypravení mělo být nasazeno jiné vozidlo. Do grafu ale chceme zanést i informaci, že došlo k určité nesrovnalosti mezi nasazenými vozidly, a tak je připsán i záznam o chybě.

Poslední sloupec type je vyplněn na základě evidenčního čísla ze sloupce bus_actual, které je spárováno s příslušným typem vozidla z tabulky seznam_busu. Pokud sloupec bus_actual je „CHYBA“, typ autobusu se určí podle evidenčního čísla ze sloupce apc_evč, jelikož opět, toto vozidlo muselo být v provozu, když k němu v souboru Vypravení vozů s APC existuje záznam (viz obrázek 24).

	v_datum	v_linka	v_poradi	apc_linkmen	apc_poradi	v_vuz	apc_evč	original_order	bus_actual	melo_apc	type
203	2023-02-01	135	14	135	14	6669	6978	203	CHYBA	A_CHYBA	NB 18

Obrázek 24 - Ukázka záznamu s hodnotou CHYBA (v provozu více než jedno vozidlo)

Zpravidla dojde ke změně vozidla v případě, že dojde k chybě v komunikaci s původně nasazeným vozidlem. To, že identifikační číslo vozidla, které bylo plánované nasadit, se neshoduje s číslem vozidla, které pak skutečně vysílalo signál, je následně v grafu zaznačeno žlutým podkladem. Není k dispozici časový záznam, a tak není jisté, jestli obě vozidla byla skutečně v provozu, nebo zda se na chybu v plánovaném vozidle přišlo ještě před vyjetím z garáže a v provozu tak bylo jen náhradní vozidlo. Každopádně je jisté, že plánované vozidlo systém APC nemělo. V grafu je zaneseno pouze to náhradní vozidlo, jelikož neexistuje duplicitní řádek se stejným datem, linkou a pořadím (není možné, že obě vozidla měla systém APC). Z dat nelze s jistotou říct, zda bylo plánované vozidlo v provozu, a tak v grafu není zaneseno.

Stejně tak se může stát, že na jedno pořadí byla nasazena dvě vozidla a obě byla vybavena systémem APC. To se dozvíme ze souboru Vypravení vozů s APC, protože obsahuje vysílání ze dvou různých vozidel pro ten samý oběh. Jedno z těchto vozidel se může shodovat s plánovaným vozidlem, to druhé pak bude logicky odlišné od plánovaného (viz obrázek 25).

	v_datum	v_linka	v_poradi	apc_linkmen	apc_poradi	v_vuz	apc_evč	original_order	bus_actual	melo_apc	type
244	2023-03-29	137	1	137	1	3999	3944	23208	CHYBA	A_CHYBA	NB 12
245	2023-03-29	137	1	137	1	3999	3999	23209	3999	A	NB 12

Obrázek 25 - Ukázka záznamu, kdy byla nasazena dvě vozidla vybavena systémem APC, jedno z vozidel (apc_evč) se shoduje s plánovaným vozidlem (v_vuz)

Také se ale nemusí s plánovaným vozidlem shodovat ani jedno z dvou skutečně nasazených vozidel, pak víme, že plánované vozidlo vybaveno nebylo. Opět nelze určit, jestli i toto plánované vozidlo bylo v provozu či nikoliv.

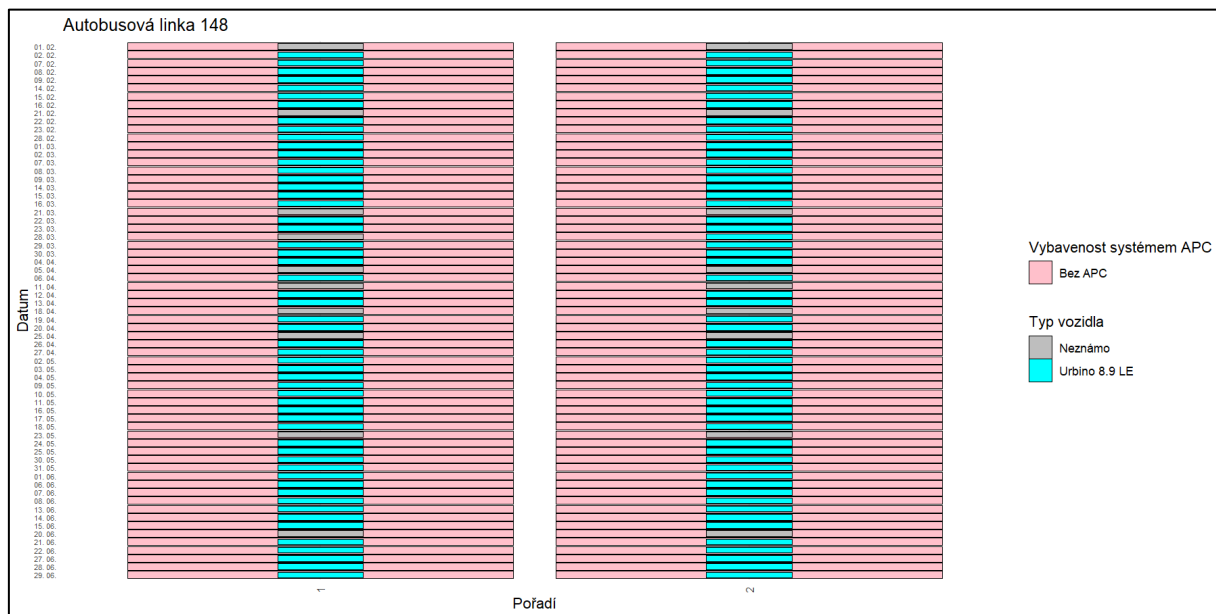
	v_datum	v_linka	v_poradi	apc_linkmen	apc_poradi	v_vuz	apc_evč	original_order	bus_actual	melo_apc	type
446	2023-05-16	137	1	137	1	3724	4046	43398	CHYBA	A_CHYBA	NB 12
447	2023-05-16	137	1	137	1	3724	5034	43399	CHYBA	A_CHYBA	NB 18

Obrázek 26 - Ukázka záznamu, kdy byla nasazena minimálně dvě vozidla, dvě s jistotou vybavená systémem APC, ani jedno z nich (apc_evč) se neshoduje s plánovaným vozidlem (v_vuz)

Za celé sledované období nedošlo k situaci, že by se na provozu jednoho pořadí podílela více než tři vozidla. Na vybraných linkách nyní prakticky ukážu situace, které byly popsány teoreticky.

Linka 148 (0 %)

Pokrytí jednotlivých linek systémem APC úzce souvisí s typy autobusů, které bývají obvykle na dané lince vypraveny. Linka 148 (I.P.Pavlova – Karlovo Náměstí – I.P.Pavlova) má výsledný pokrytí 0 %. Rozložení pokrytí na jednotlivé oběhy pro tuto linku je zobrazeno v 2D grafu níže (viz Skript 6).

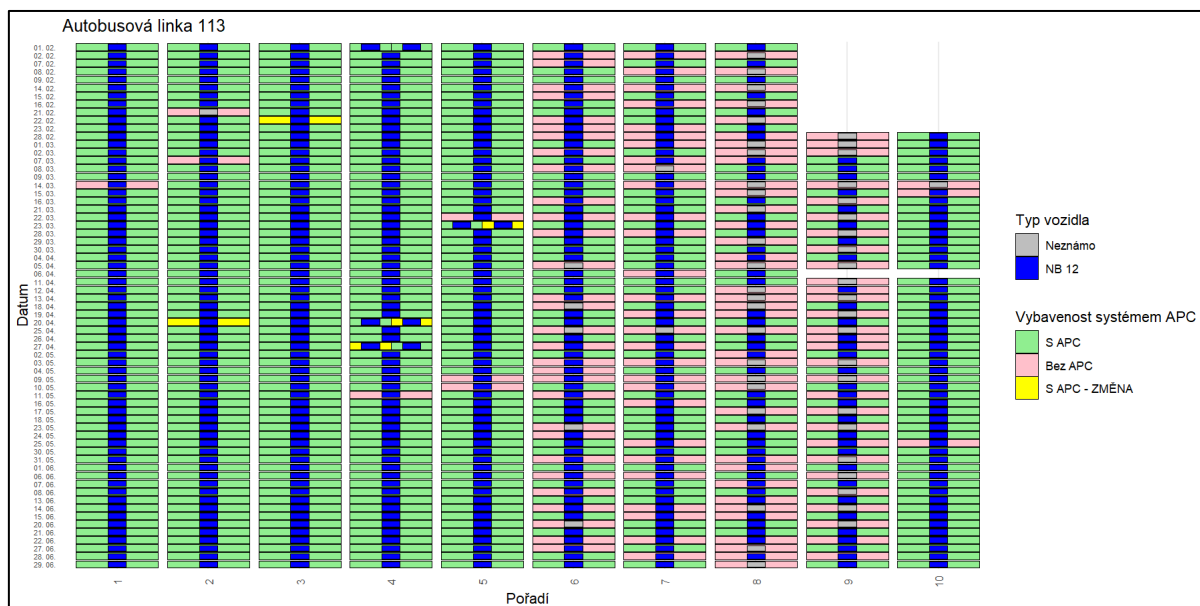


Obrázek 27 - Zobrazení pokrytí a typů nasazených autobusů na lince 148

Z grafu vidíme, že linka je v provozu po celé sledované období v podobě dvou pořadí. Na těchto dvou pořadích se střídají vozidla typu Urbino 8.9 LE. Tato vozidla dle předchozí analýzy zařízením APC nedisponují a tento graf skutečnost potvrzuje. Za sledované období se na lince vystřídal minimálně 11 různých vozidel typu Urbino 8.9 LE (zjištěno dotazem, viz Dotaz 11) a žádné z nich APC zařízení nemělo.

Linka 113 (78,69 %)

Fenomén, kdy dojde k výměně jednoho vozidla za jiné beze změny pořadí, se objevuje skrze celé sledované období.



Obrázek 28 - Zobrazení pokrytí a typů nasazených autobusů na lince 113

V grafu pokrytí jednotlivých pořadí linky 113 vidíme u pořadí 4 (resp. 5) změnu. Znamená to, že pro dny 1. 2., 20. 4. a 27. 4. bylo 4. pořadí linky 113 operováno dvěma vozidly. Z dat, která mi jsou dostupná, nelze určit pořadí nasazení těchto vozidel. Lze předpokládat, že původní vozidlo přestalo vysílat signál a bylo nahrazeno novým, plně funkčním vozidlem. Víme, že obě vozidla byla vybavena systémem APC.

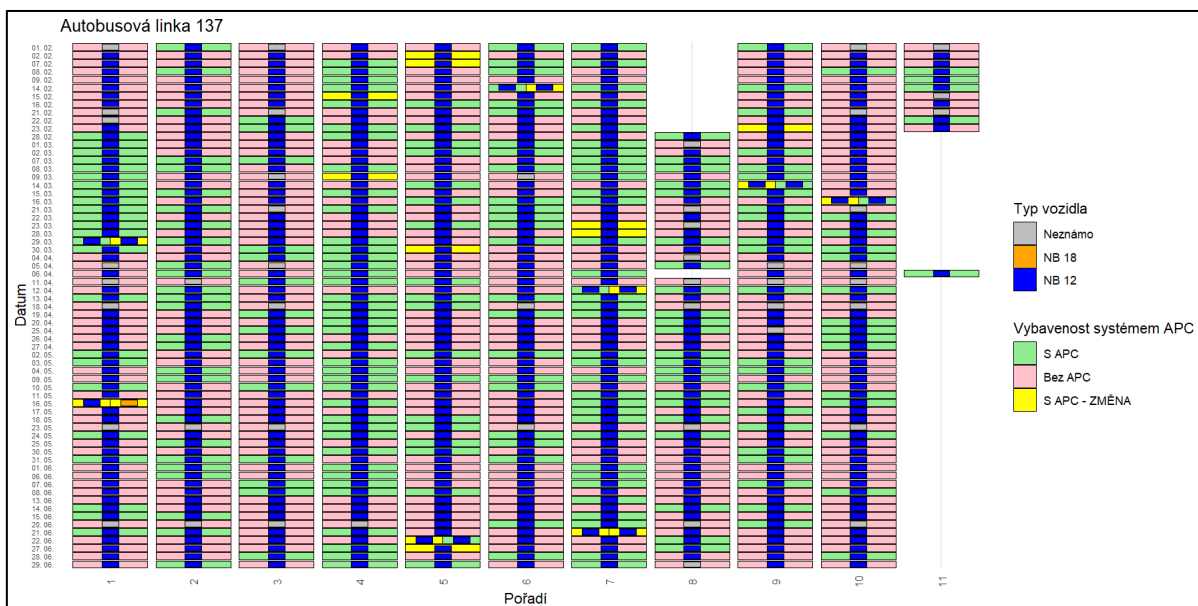
V případě, že je pořadí obslouženo dvěma vozidly, a obě mají zelený podklad (4. pořadí, den 1. 2.), pak obě vozidla byla vybavena systémem APC a nemáme záznam o tom, jaké vozidlo bylo plánované, a jestli bylo též vypraveno.

	v_datum	v_linka	v_poradi	apc_linkmen	apc_poradi	v_vuz	apc_evč	bus_actual	melo_apc	type
4	2023-02-01	113	4	113	4	NULL	3914	3914 A		NB 12
5	2023-02-01	113	4	113	4	NULL	4044	4044 A		NB 12

Obrázek 29 - Záznam ze zdrojové tabulky pro zobrazení, 4. pořadí dne 1.2.2023

Linka 137 (39,54 %)

V případě linky 137 je vidět, že provoz dvou vozidel v rámci jednoho pořadí není až tak nestandardní událost.



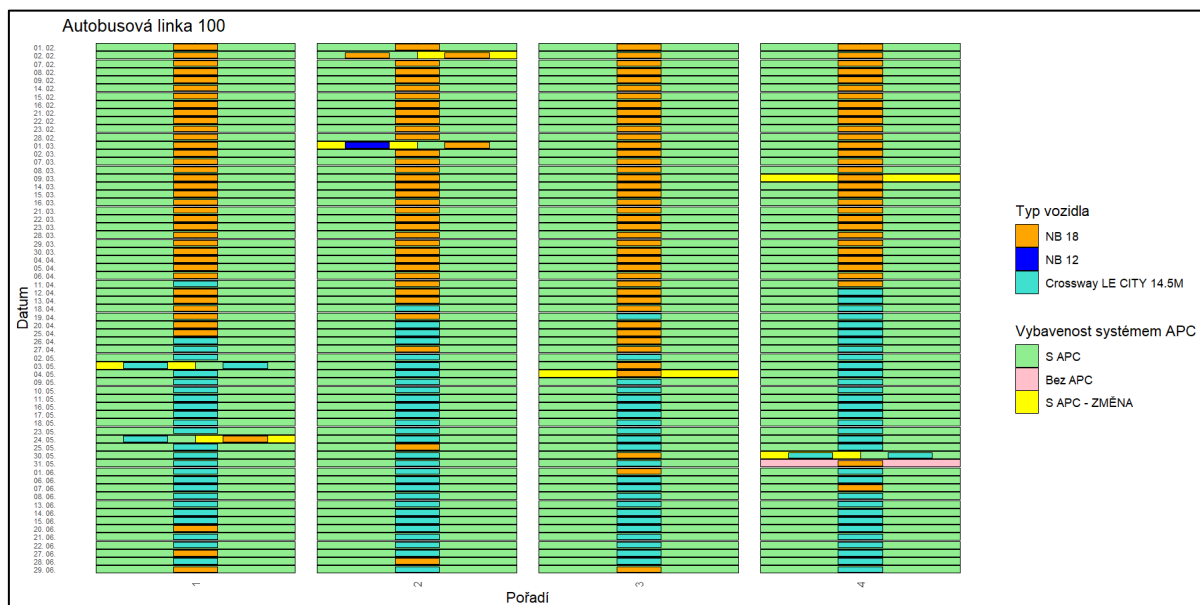
Obrázek 30 - Zobrazení pokrytí a typů nasazených autobusů na lince 137

V některých případech je plánované vozidlo (zelený podklad) vyměněno za jiné, které nebylo v plánu použito (žlutý podklad). Sled dvou vozidel v jeden den na jednom pořadí je zobrazen na základě vzestupného evidenčního čísla vozidel, nikoli na základě časové posloupnosti. Z toho důvodu je někdy první vůz podbarven zeleně a druhý žlutě (1. pořadí, den 29. 3.) a někdy naopak (9. pořadí, 14. 3.). Můžeme se tedy domnívat, že nejdříve byl nasazen plánovaný vůz, ale z důvodu nefunkční komunikace byl namísto něj v průběhu dne nasazen neplánovaný vůz.

V jiných případech je neplánovaným vozidlem (žlutý podklad) nahrazeno jiné vozidlo, které rovněž nebylo v plánu použito (také žlutý podklad). To se stalo například dne 16. 5. na prvním pořadí. Zde byl změněn též typ vypraveného vozu. V této situaci (viz obrázek 26) s jistotou byla v provozu dvě vozidla, 4046 a 5034, a obě byla vybavena zařízením APC. V plánu bylo nasadit vozidlo 3724, a jelikož toto vozidlo systém APC nemělo, nevíme, zda bylo nějakou dobu v provozu, a nebo vůbec nevyjelo. Proto jsou v grafu zobrazena pouze dvě náhradní vozidla.

Linka 100 (99,62 %)

U linky 100 je možno pozorovat změnu vozidel pravidelně posílaných na tuto linku.



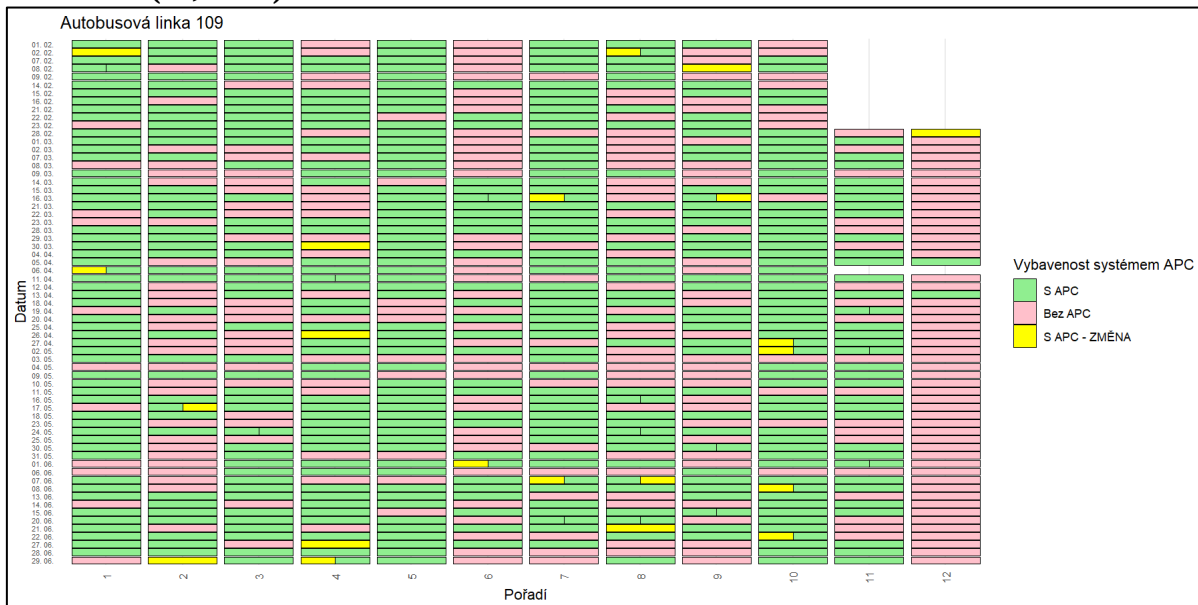
Obrázek 31 - Zobrazení pokrytí a typů nasazených autobusů na lince 100

Přibližně do poloviny dubna byla linka obsluhována vozidly typu NB 18, od poloviny dubna dál potom převážně vozidly typu Crossway LE CITY 14.5M. Důvodem může být nejspíš pravidelně nevyužitá kapacita kloubového autobusu a potřeba jej nasazovat na více vytížených linkách. Vozidla typu Crossway LE CITY 14.5M jsou uzpůsobena pro pohodlný převoz zavazadel. Jelikož se jedná o linku na letiště (Zličín–Letiště), kapacita vozidel mohla být často využívána právě zavazadly, nikoli osobami. Vozidlo se speciálním prostorem pro zavazadla tak umožnilo nasazovat kratší vozidla, než jaká jezdila na lince dříve.

3.5 Pokrytí oběhů systémem APC během dne

Předpoklad zní, že vozidla vybavená systémem APC jsou přednostně posílána na ranní a odpolední oběhy („šejdry“), a oběhy, které jsou vypravovány na celý den („túráky“), vybavovány nejsou. To by znamenalo, že by bylo možné sledovat trend hustěji pokrytých časných a pozdních pořadí a méně pokrytých pořadí uprostřed dne. Pro tuto analýzu bude vhodné se podívat na linky, které mají přibližně 50-70 % pokrytí systémem APC, aby bylo možné rozpoznat případný vzorec v nasazování vybavených vozidel. Graf je opět zobrazen v rámci programu RStudio (viz Skript 7).

Linka 109 (62,95 %)



Obrázek 32 - Zobrazení pokrytí pořadí na lince 109

Linka 191 (66,41 %)



Obrázek 33 - Zobrazení pokrytí pořadí na lince 191

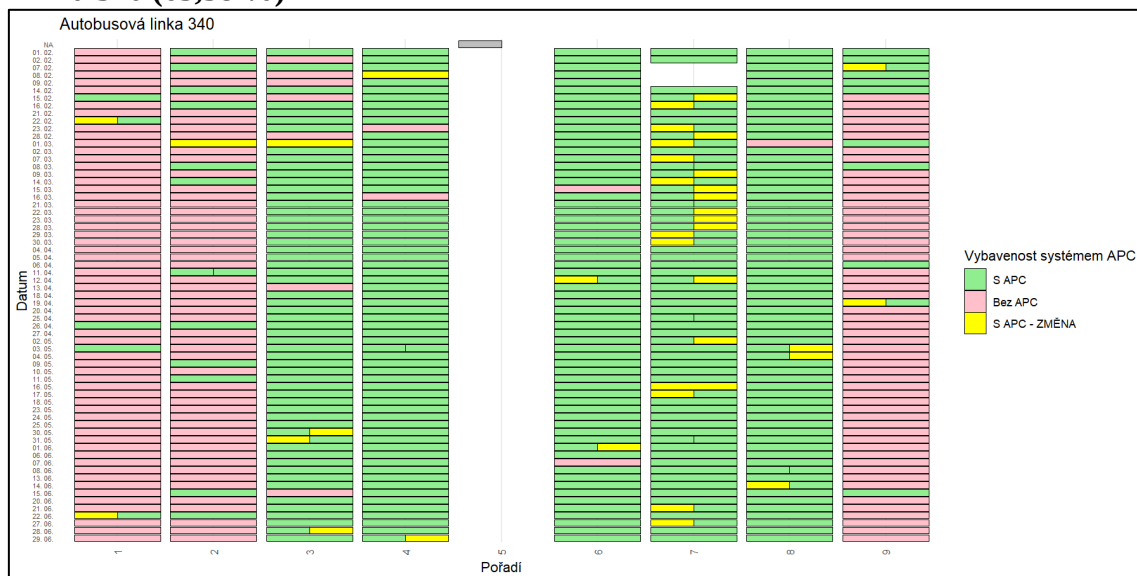
U linek 109 a 191 lze rozhodně vidět trend v pravidelně se opakujícím pokrytí určitých pořadí. U linky 109 poslední pořadí (číslo 12) bylo ve sledovaném období jen velmi málo pokryto. Naopak u linky 191 je poslední pořadí (číslo 18) pravidelně pokryté. Jinak se u obou linek zdá, že pokrytí pořadí v jednotlivé dny je realizováno náhodně.

Linka 113 (78,69 %)



Obrázek 34 - Zobrazení pokrytí pořadí na lince 113

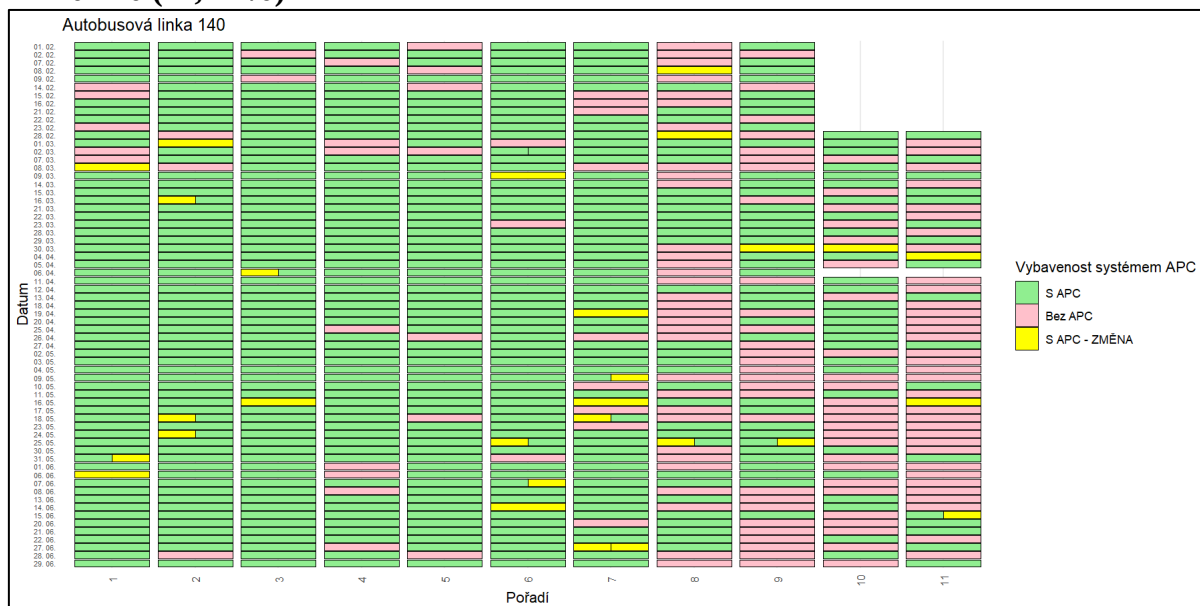
Linka 340 (65,58 %)



Obrázek 35 - Zobrazení pokrytí pořadí na lince 340

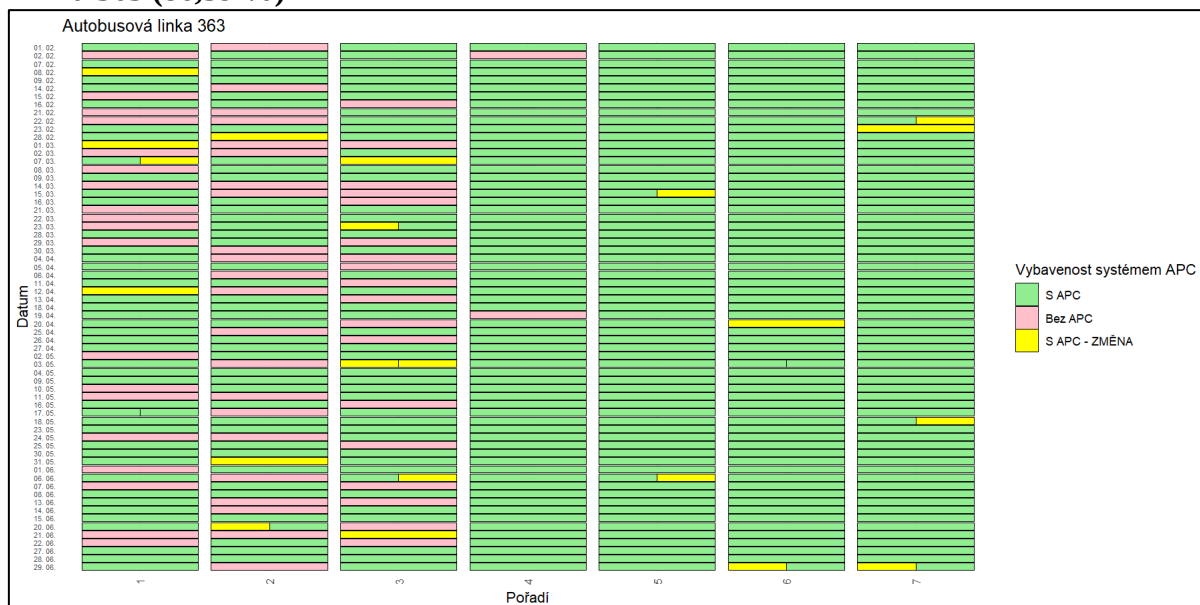
Dalšími vybranými linkami jsou linky 113 a 340. U obou linek je vidět pravidelnost v nasazování vybavených vozidel, ale opačným způsobem. Linka 113 mívá obvykle pokrytá krajní pořadí, zatímco pořadí uprostřed pokryta nejsou. U linky 340 je trend přesně opačný, tedy krajní pořadí pokryta nejsou, zatímco pořadí uprostřed ano (5. pořadí zde chybí, dle jízdního řádu nebylo po celé sledované období vypraveno).

Linka 140 (77,47 %)



Obrázek 36 - Zobrazení pokrytí pořadí na lince 140

Linka 363 (86,59 %)



Obrázek 37 - Zobrazení pokrytí pořadí na lince 363

U linek 140 a 363 je znát další, odlišný vzorec v nasazování pokrytých spojů. Linka 140 pravidelně měla pokryta pořadí z kraje dne, pozdější pořadí bývala pokrytá méně. Naopak u linky 363 docházelo k vypravování vybavených vozů u pořadí s vyšším číslem vypravovaných později. Ranní pořadí byla pokryta méně.

Výsledkem této analýzy je závěr, že nelze jednoznačně stanovit trend, který by sledovaly všechny linky. Nelze tedy tvrdit, že vozidla vybavená systémem APC jsou přednostně posílána na ranní a odpolední oběhy.

4. Chování parametru obsazenosti linky v čase

4.1 Obecný přístup k analýze

Byly vybrány čtyři linky, jejichž obsazenost je sledována v čase. Pro spolehlivé výsledky je žádoucí vybrat takovou linku, která má procento pokrytí alespoň 50 %. Vybrány byly linky:

Tabulka 6 - Soupis blíže zkoumaných linek

Číslo linky	Celkové pokrytí APC	Průměrná doba jízdy	Délka linky	Trasa	Počet zastávek	Počet pořadí za den
	[%]	[min]	[km]			
119	81,63	17	8	Nádraží Veleslavín – Letiště / Terminál 1	12	17
133	64,10	20	6	Florenc – Sídliště Malešice	16	9
201	93,64	56	22	Nádraží Holešovice – Černý Most	42	22
352	94,62	20	7	Praha, Luka – Jinočany, náměstí	15	10

Pro zobrazení obsazenosti linky v čase slouží 3D graf. Zdrojovým souborem pro vykreslení grafu jsou data o vypravení vozů vybavených systémem APC poskytnuta společností ROPID v podobě CSV souboru (apc-2023-02-06) – Vypravení vozu s APC. Na ose y je zobrazen čas, na ose x jednotlivé zastávky linky a v každém průsečíku času a zastávky je pole vybarveno barvou. Dle legendy barva odpovídá počtu cestujících, kteří se nacházeli ve vozidle po odjezdu z dané zastávky.

Čas odjezdu zobrazen na ose y odpovídá údajům ze sloupce s názvem casjr, graf tedy nebere v potaz zpoždění spojů a řídí se předpokládaným odjezdem daného pořadí ze zastávky. Záznam o počtu cestujících opouštějících danou zastávku ve vozidle linky autobusu je získán z příslušného záznamu ve sloupci s názvem cestujících.

Posloupnost zastávek na ose x je určena podle standardně se vykazujícího pořadí, jehož číslo je určeno ze zdrojového CSV souboru. Zastávky jsou zobrazeny z první konečné zastávky, odkud spoj vyjel, přes druhou konečnou zastávku a zpět do první konečné zastávky. Lze tedy sledovat obsazenost spoje po celou jeho trasu, nikoli pouze v jednom směru. Zároveň tím, že je vybráno pořadí, u kterého je zkontrolováno, že jelo standardní trasou, je zajištěno, že se v grafu neukážou zastávky navíc, například kvůli výluce či odklonu linky jinou trasou. Pokud se výjimečně takovéto zastávky v souboru pro danou linku objevily, je ošetřeno, aby se v grafu vůbec neobjevily.

Hodnoty počtu cestujících reprezentované v grafu příslušnou barvou jsou průměrnou hodnotou vypočítanou za celé sledované období (únor-červen) ze všech záznamů opakujících se ve stejném čase dle jízdního řádu. Zároveň jsou přenášeny odpovídajícím procentem pokrytí pro danou linku. Logika zobrazení dat v grafu je následující. Jsou upraveny původní počty cestujících tak, aby bylo zohledněno procento pokrytí, a poté je vypočítán průměr pro každou kombinaci zastávky a času. Tím je pro danou kombinaci zastávky a času získána hodnota vypovídající za celé sledované období. Vzorec pro výpočet uveden zde:

$$prumer_cestujících = \frac{1}{n} \sum_{i=1}^n \left(\frac{cestujících_i \cdot 100}{pokrytí} \right)$$

Vztah 1 - Výpočet průměrného počtu cestujících ve vozidle pro danou kombinaci zastávky a času

Kde:

- *prumer_cestujících* je výsledná hodnota průměrného počtu cestujících pro danou kombinaci zastávky a času
- *n* je počet datových záznamů pro danou kombinaci zastávky a času
- *cestujících_i* je naměřený počet cestujících ve vozidle pro i-tý datový záznam
- *pokrytí* je procento pokrytí pro danou autobusovou linku

Kód pro vytvoření grafu obsazenosti linky v čase byl vytvořen v programu RStudio (viz Skript 8).

U každé linky je dále vybrána jedna zastávka a je sledován průměrný počet nastupujících cestujících v této zastávce do dané linky v obou směrech. Vybrány byly zastávky:

Tabulka 7 - Soupis blíže zkoumaných zastávek

Zastávka	Linka	Pokrytí	Uzel	Zast	Směr
		[%]			
Navigátorů	119	81,63	80	1	Letiště / Terminál 1
				2	Nádraží Veleslavín
Pernerova	133	64,10	3018	1	Sídliště Malešice
				2	Florenc
Rajská zahrada	201	93,64	818	1	Černý Most
				2	Nádraží Holešovice
Amforová	352	94,62	1029	1	Praha, Luka
				2	Jinočany, náměstí

Zdrojovým souborem pro tento 2D graf je stejně jako u předchozího grafu soubor Vypravení vozů s APC. Graf zobrazuje na ose y průměrný počet cestujících nastupujících v dané zastávce do dané linky, na ose x čas. Zobrazeny jsou oba směry, jsou odlišeny barvou a řádně popsány legendou.

Čas je opět přejat ze sloupce s názvem casjr, zpoždění spojů tedy není do grafu zahrnuto. Zobrazený počet cestujících je záznamu odpovídající hodnota ve sloupci s názvem nastup.

Stejně jako předchozí graf, i tento graf zobrazuje zprůměrované hodnoty za celé sledované období. Ty jsou získány obdobně a to dle následujícího vzorce.

$$prumer_nastup = \frac{1}{n} \sum_{i=1}^n \left(\frac{nastup_i \cdot 100}{pokryti} \right)$$

Vztah 2 - Výpočet průměrného počtu nastupujících cestujících v čase pro jednu zastávku

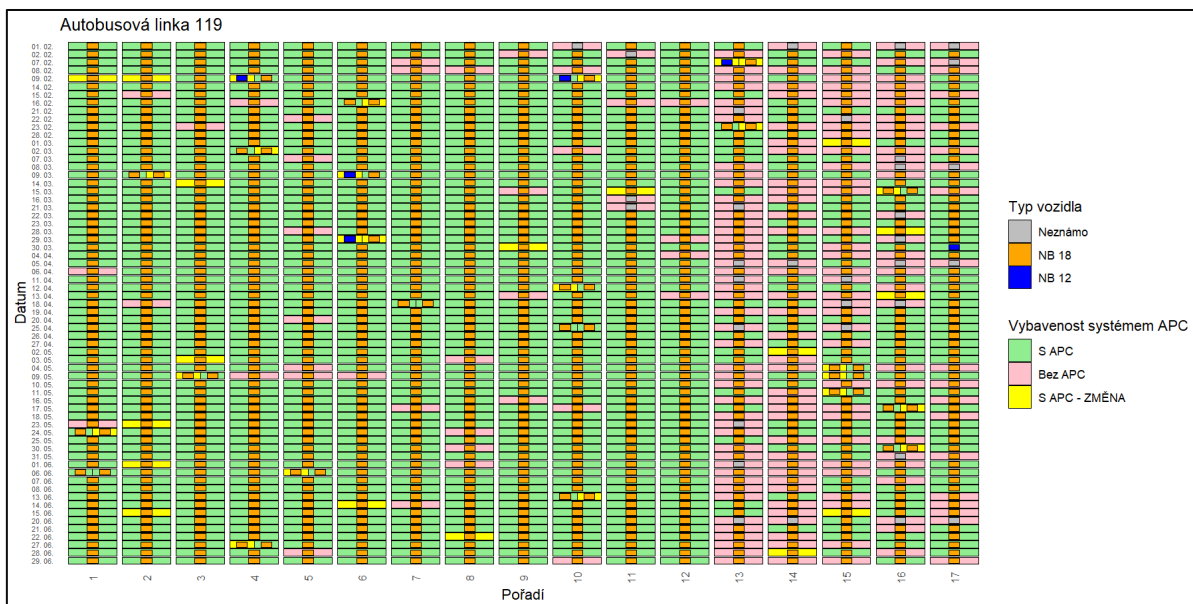
Kde:

- *prumer_nastup* je výsledná hodnota průměrného počtu nastupujících cestujících v zastávce pro daný čas
- *n* je počet datových záznamů pro konkrétní čas
- *nastup_i* je naměřený počet nastupujících cestujících pro i-tý datový záznam
- *pokryti* je procento pokrytí pro danou autobusovou linku

Kód tvořící graf zobrazující průměrný počet nastupujících cestujících v zastávce do spojů dané linky je rovněž napsán v program RStudio (viz Skript 9).

4.2 Linka 119

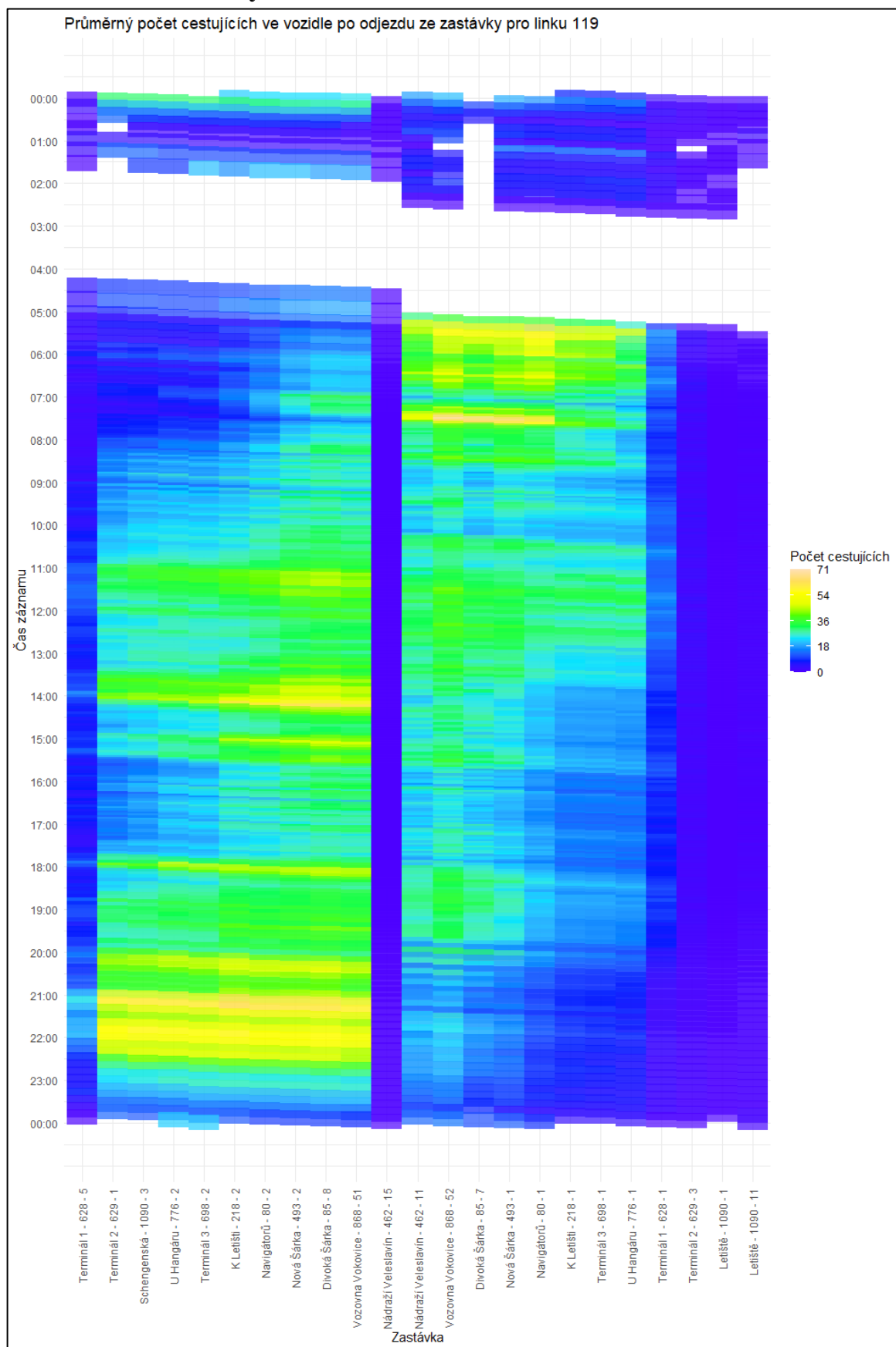
Na lince 119 (Nádraží Veleslavín – Letiště / Terminál 1) byla převážně vypravována vozidla typu NB 18. V grafu níže lze vidět, že pravidelně byla spíše pokryta dřívější pořadí.



Obrázek 38 - Zobrazení pokrytí a typů nasazených autobusů na lince 119

Celkové pokrytí linky systémem APC je 81,63 %. Počty cestujících zobrazované v následujících grafech jsou vynásobeny hodnotou $\frac{100}{81,63}$, aby co nejvíce odpovídaly skutečnému stavu.

4.2.1 Obsazenost linky 119 v čase



Obrázek 39 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 119

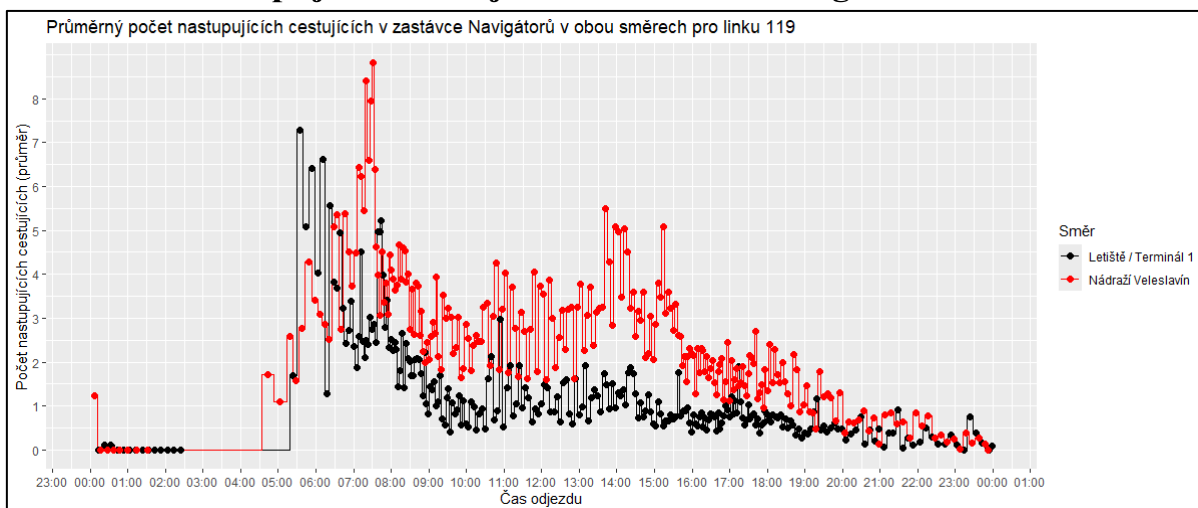
Graf obsazenosti vozidel linky 119 zobrazuje tři hlavní dopravní špičky. První ve směru do zastávky Letiště v ranních hodinách, v časovém rozmezí přibližně 5:00-7:30. Je možno předpokládat, že se jedná o cestující, kteří míří na letiště za prací. Stejně tak se může jednat o cestující mířící na letiště za účelem odletu. Dopolední odlet je cestujícími žádaný a vzhledem k potřebě být na letišti s dostatečnou časovou rezervou by cesta kolem 6. hodiny ranní odpovídala tomuto požadavku [22]. Důvodem vyššího vytížení zastávky K Letišti (218 / 1, 2) může být fakt, že v její blízkosti se nachází obchodní centrum.

Druhá, ne tolik koncentrovaná vlna cestujících je vidět v časovém rozmezí 11:00-15:30 ve směru do zastávky Nádraží Veveř. Toto období by se standardně nazvalo poledním sedlem. Jelikož se ale jedná o linku zajišťující obsluhu letiště, nedá se předpokládat, že úbytek poptávky po přepravě nastane ve stejný čas, jako by tomu bylo u běžné městské linky. Lze vidět, že je tomu právě naopak. Důvodem může být atraktivnost příletu v této denní době – přes den, za světla, mimo místní dopravní špičku, ideální pro odpolední ubytování v cílové destinaci [22].

Třetí významná vlna se odehrává v časovém rozmezí 18:00-22:30 ve směru do zastávky Nádraží Veveř. Může se jednat jak o zaměstnance letiště vracející se z práce, tak o cestující, kteří právě přiletěli.

Obecně se většina nástupů ve směru do zastávky Nádraží Veveř odehrává v zastávce Terminál 2 (629/1). Většina nástupů ve směru do zastávky Letiště je realizována v zastávce Nádraží Veveř (462/11) a v zastávce Vozovna Vokovice (868/52). Ne však v takové míře, jako je tomu u zastávky Terminál 2. Shrnutí grafu tedy vypovídá o tom, že veřejnou hromadnou dopravu, konkrétně linku 119, využívali cestující především k dopravě z letiště oproti dopravě na letiště.

4.2.2 Počet nastupujících cestujících v zastávce Navigátorů v čase

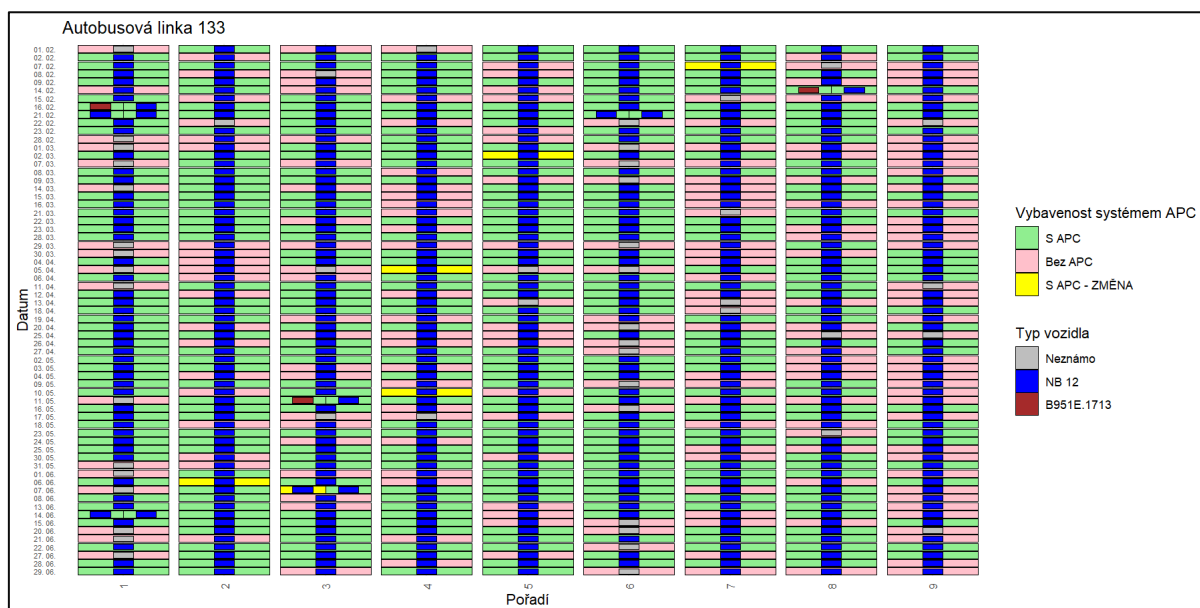


Obrázek 40 - Průměrný počet nastupujících cestujících v zastávce Navigátorů

Obecně více nástupů bylo v zastávce Navigátorů (80 / 1, 2) realizováno ve směru do zastávky Nádraží Veveř. Zastávka Navigátorů se nachází na výjezdu z Prahy v ulici Evropská. Rozsah průměrů nastupujících cestujících se pohybuje od 0 do 9 cestujících. V okolí zastávky se nachází nízkopodlažní zástavba na jedné straně, sad Zlodějka a park na straně druhé. V přímé blízkosti se nenachází žádný speciální zdroj ani cíl přepravy. Odhadem tedy je, že nastupující cestující byli především místní obyvatelé s přepravními potřebami stálého charakteru.

4.3 Linka 133

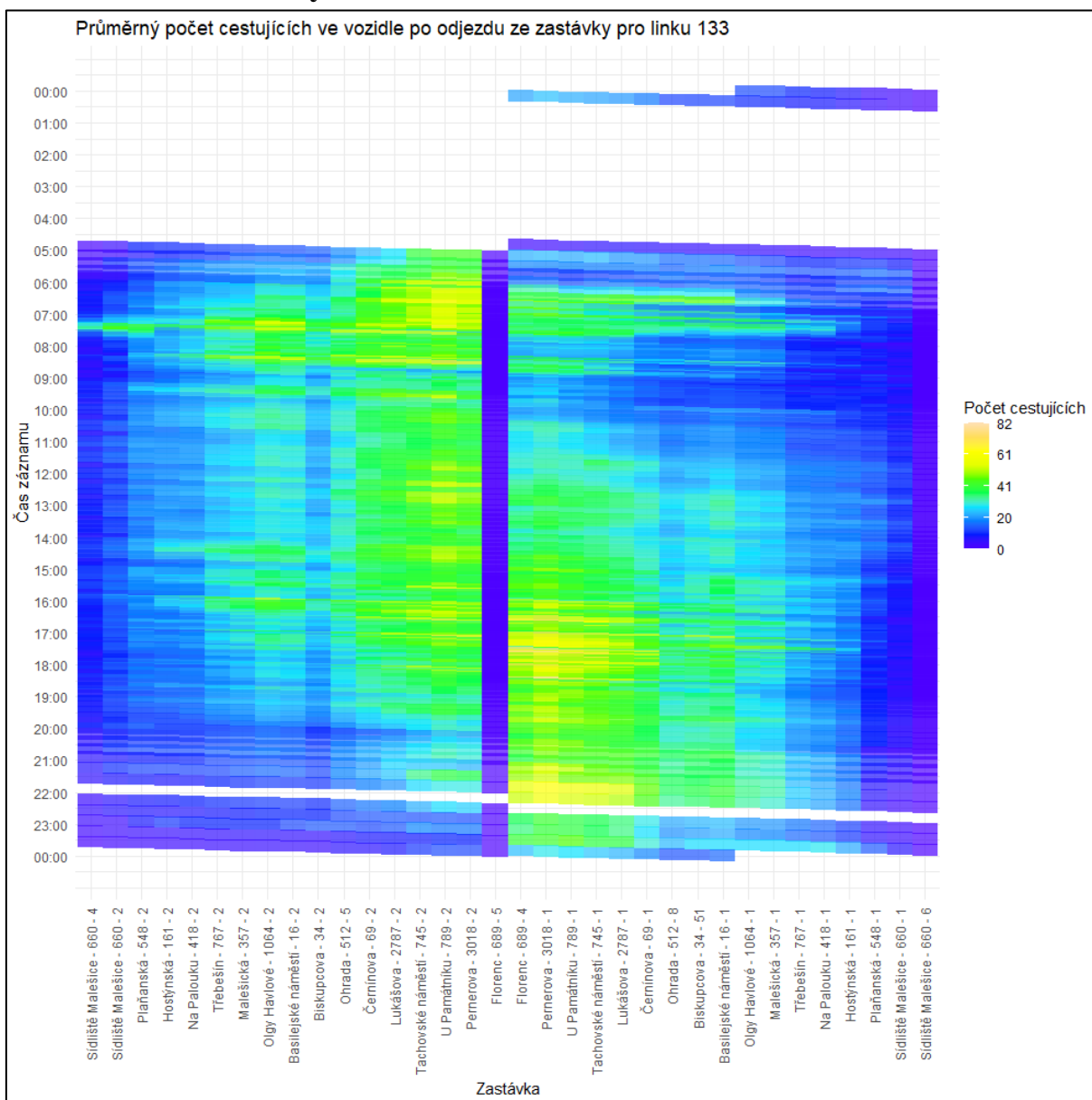
Na lince 133 (Florenc – Sídliště Malešice) byla převážně vypravována vozidla typu NB 12. V grafu níže lze pozorovat, že pokrytí pořadí skrze období byla realizována bez jasně rozpoznatelné pravidelnosti.



Obrázek 41 - Zobrazení pokrytí a typů nasazených autobusů na lince 133

Celkové pokrytí linky systémem APC je 64,10 %. Počty cestujících zobrazované v následujících grafech jsou vynásobeny hodnotou $\frac{100}{64,10}$, aby co nejvíce odpovídaly skutečnému stavu.

4.3.1 Obsazenost linky 133 v čase



Obrázek 42 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 133

Linka 133 má obslužný až sběrný charakter. Pro mnoho cestujících je ideálním spojením okrajové části města s centrem. To se projevuje významně vyšším vytížením linky v oblasti Florenc-Ohrada než v oblasti Biskupcova-Sídlíště Malešice (v obou směrech).

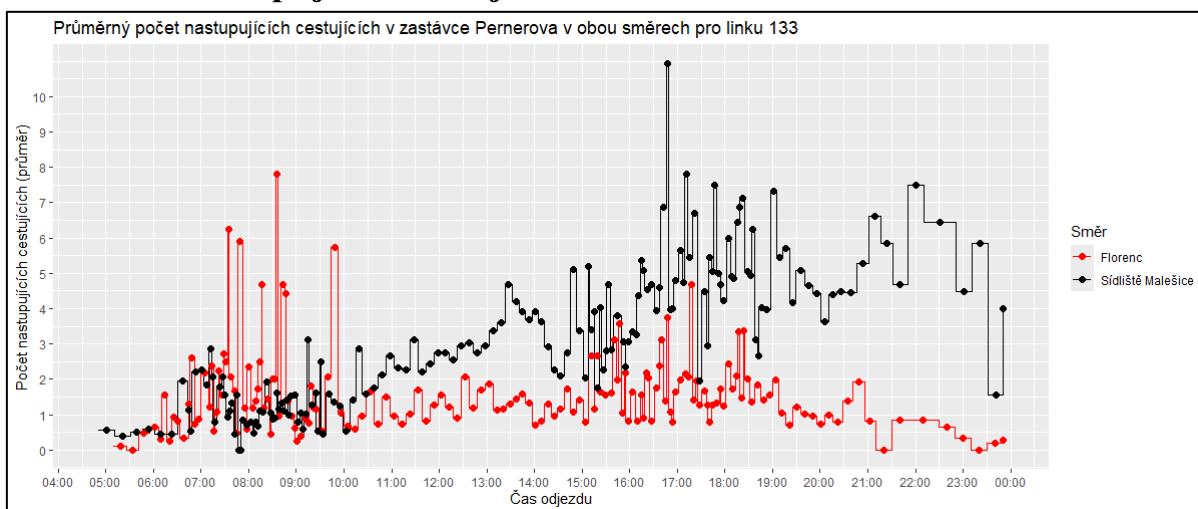
V obou směrech lze vidět standardní vývoj denního rozdělení počtu cestujících na lince. Ranní špička probíhá v časovém rozmezí 5:00-8:30, následuje zde jen lehce výrazné polední sedlo v časovém rozmezí 9:00-14:00. Odpolední špička probíhá v časovém rozmezí 15:30-18:30. To odpovídá předpokládanému provozu, kdy ráno cestující míří do práce a do škol a odpoledne se vrací zpět do výchozího místa. Zajímavá špička je též vidět v časovém rozmezí 21-22:30.

Ve směru do zastávky Florenc je polední sedlo méně výrazné než v opačném směru. To může mít za důvod fakt, že se již linka pohybuje v širším centru města, a je využívána pro krátké přesuny na blízké vzdálenosti, kdy není pro cestujícího rozhodující, kterou linku využije.

Místní cíle na trase linky jsou v zastávce Florenc (689/4,5) – přestup na metro B a C a přestup na tramvajové linky, zastávka U Památníku (789/1,2) – Armádní muzeum Žižkov nebo zastávka Sídliště Malešice (660 / 1,2,4,6), v jejíž blízkosti se nachází zámek Malešice a kaple sv. Václava.

Z grafu lze též vyčíst, že mnozí cestující vystupují v zastávce Biskupcova (34/2) ve směru do zastávky Florenc a v zastávkách Ohrada (512/8) a Třebešín (767/1) ve směru do zastávky Sídliště Malešice. Důvodem výstupu v zastávce Biskupcova může být přestup na jinou linku, například tramvajovou. V zastávce Ohrada může být cílem park Vítkov s národním památníkem a v zastávce Třebešín je důvodem výstupu především přilehlé sídliště.

4.3.2 Počet nastupujících cestujících v zastávce Pernerova v čase

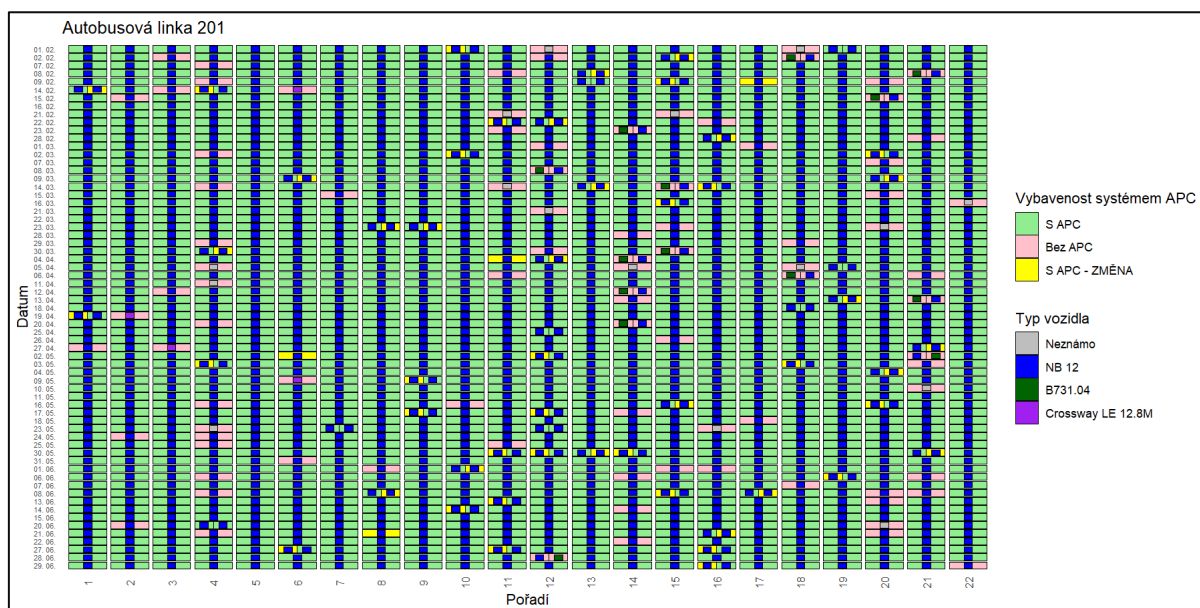


Obrázek 43 - Průměrný počet nastupujících cestujících v zastávce Pernerova

V zastávce Pernerova se rozsah průměrů cestujících nastupujících do spojů linky 133 pohybuje mezi 0 až 11 cestujícími. V ranní špičce je vidět větší vytížení spojů ve směru do zastávky Florenc, v odpolední špičce je tomu naopak. Obecně je cestujícími více vytížený směr z centra do zastávky Sídliště Malešice.

4.4 Linka 201

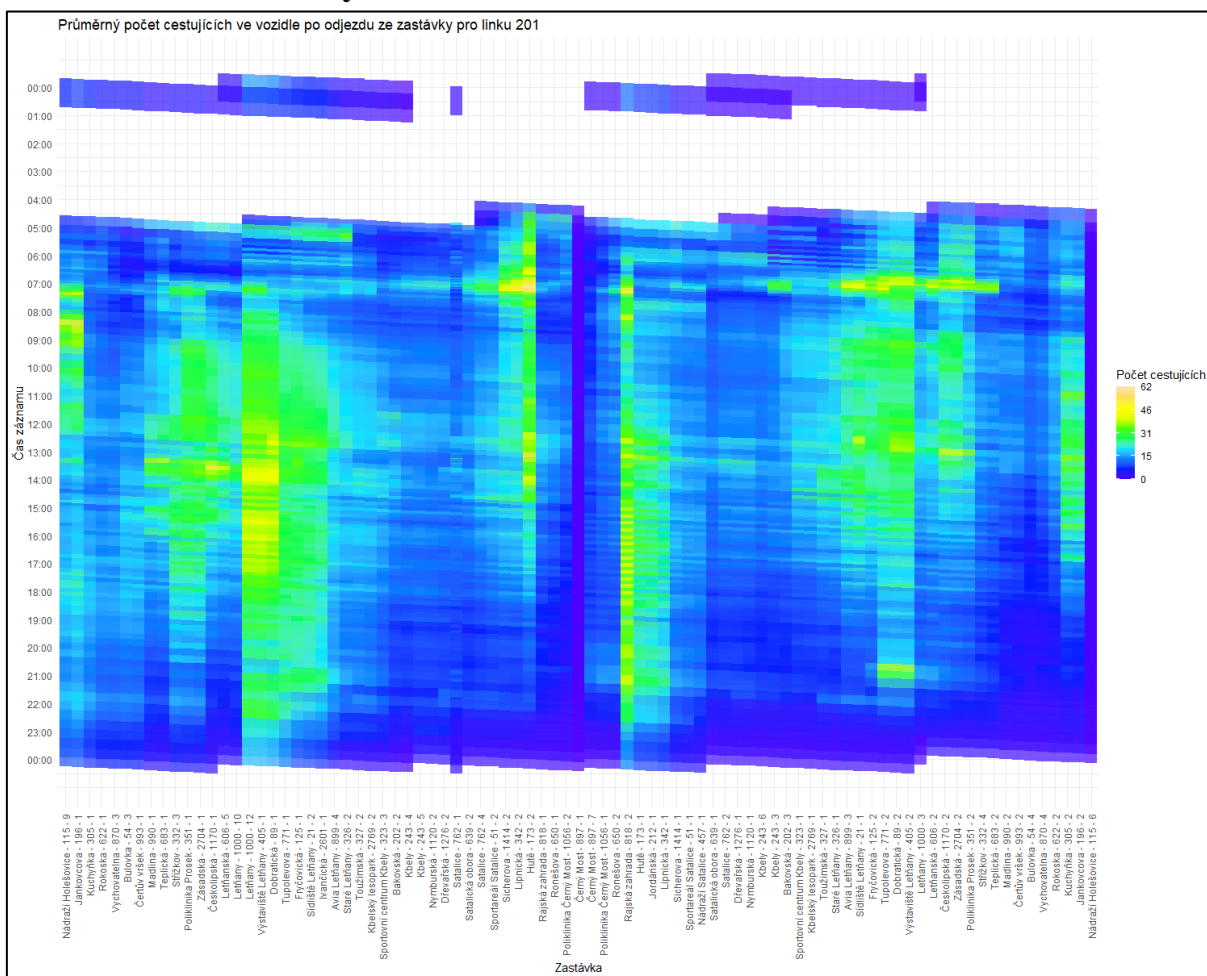
Na lince 201 (Nádraží Holešovice – Černý Most) byla převážně vypravována vozidla typu NB 12. V grafu níže lze pozorovat, že pokrytí pořadí skrze období byla realizována bez jasně rozpoznatelné pravidelnosti.



Obrázek 44 - Zobrazení pokrytí a typů nasazených autobusů na lince 201

Celkové pokrytí linky systémem APC je 93,64 %. Počty cestujících zobrazované v následujících grafech jsou vynásobeny hodnotou $\frac{100}{93,64}$, aby co nejvíce odpovídaly skutečnému stavu.

4.4.1 Obsazenost linky 201 v čase



Obrázek 45 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 201

Linka 201 je jednou z nejdelších linek provozovaných DPP co se týká počtu zastávek. Linka má metropolitní charakter a zajišťuje kapacitní obsluhu pro severovýchod Prahy. V několika zastávkách na trase umožňuje přestup na metro B a C.

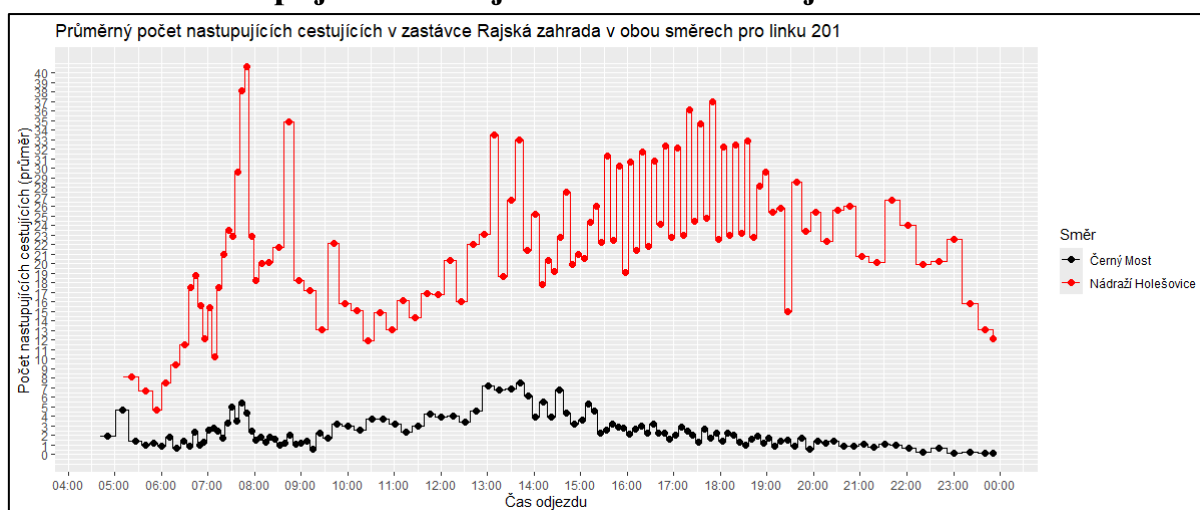
V obsazenosti spojů lze v grafu vidět závislost spíše na geografické poloze zastávky než na čase. Ve vytížených zastávkách lze pozorovat variaci dopravy v podobě ranních a odpoledních špiček, nicméně tyto špičky nejsou zvláště výrazné. Jsou jimi například zastávky Nádraží Holešovice (115/9) a Jankovcova (196/1) ve směru do zastávky Černý Most.

Naopak některé zastávky místo standardního poledního sedla jsou v časovém rozmezí 10:00-14:00 vytíženy více než v jiné denní době. Jsou jimi například zastávky v úseku Letňany (1000/12) – Ivančická (2601/1) ve směru do zastávky Černý Most a zastávky v úseku Avia Letňany (899/3) – Výstaviště Letňany (405/2) ve směru do zastávky Nádraží Holešovice. Důvodem mohou být cestující mířící do areálu výstaviště, kteří tuto návštěvu realizují v rámci standardní pracovní doby.

Specifický průběh obsazenosti spojů lze též zaznamenat v okolí zastávek s přestupem na metro jakožto páteřní linku. Zde konkrétně bude návaznost obsazenosti na denní dobu vidět nejméně, protože cestující míří na metro a z metra po celý den. Jde o fenomén, kdy se postupně cestujícími plní kapacita spoje před zastávkou s přestupem na metro, na této zastávce cestující vystoupí a dál jede spoj významně prázdnější a nebo do něj opět cestující mířící z metra nastoupí.

Do zastávky Letňany (1000/10) směrem do zastávky Černý Most jede spoj relativně plný, v této zastávce cestující přestoupí na metro. Spoj přejede do zastávky Letňany (1000/12) a noví cestující nastoupí. Odsud jede spoj opět plný a postupně se jeho obsazenost snižuje. Stejně tak cestující vystupují v zastávce Rajska zahrada (818/1) směr do zastávky Černý most a nastupují v zastávce Rajska zahrada (818/2) směr do zastávky Nádraží Holešovice.

4.4.2 Počet nastupujících cestujících v zastávce Rajska zahrada v čase



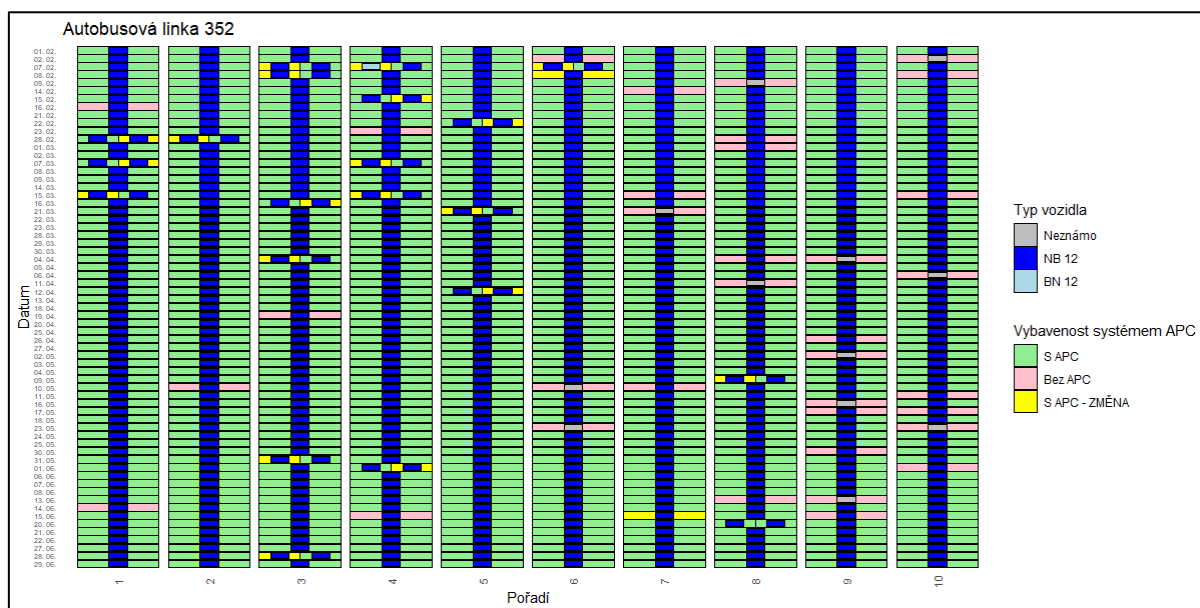
Obrázek 46 - Průměrný počet nastupujících cestujících v zastávce Rajska zahrada

Průměrný počet nastupujících cestujících v zastávce Rajska zahrada do spojů linky 201 je velice úzce spjat s tím, že je v zastávce realizován přestup na metro. Graf ukazuje, že je velký rozdíl v počtu cestujících v závislosti na směru. Ve směru do zastávky Nádraží Holešovice se maximální průměrné hodnoty počtu cestujících pohybují až okolo 40 cestujících. V opačném směru do zastávky Černý Most maximální průměrné hodnoty počtu nastupujících cestujících nepřekročily číslo 8.

Obě zastávky Rajska zahrada i Černý Most skýtají možnost přestupu na metro B. Mezi nimi leží 2 zastávky. Ti cestující, kteří nastupují do linky 201 v zastávce Rajska zahrada ve směru do zastávky Černý Most, tedy pravděpodobně mají cíl cesty v jedné z následujících dvou zastávek, a takových cestujících není v poměru k celku tolik. To je pravděpodobně důvodem, proč se počty nastupujících cestujících v obou směrech tolik liší.

4.5 Linka 352

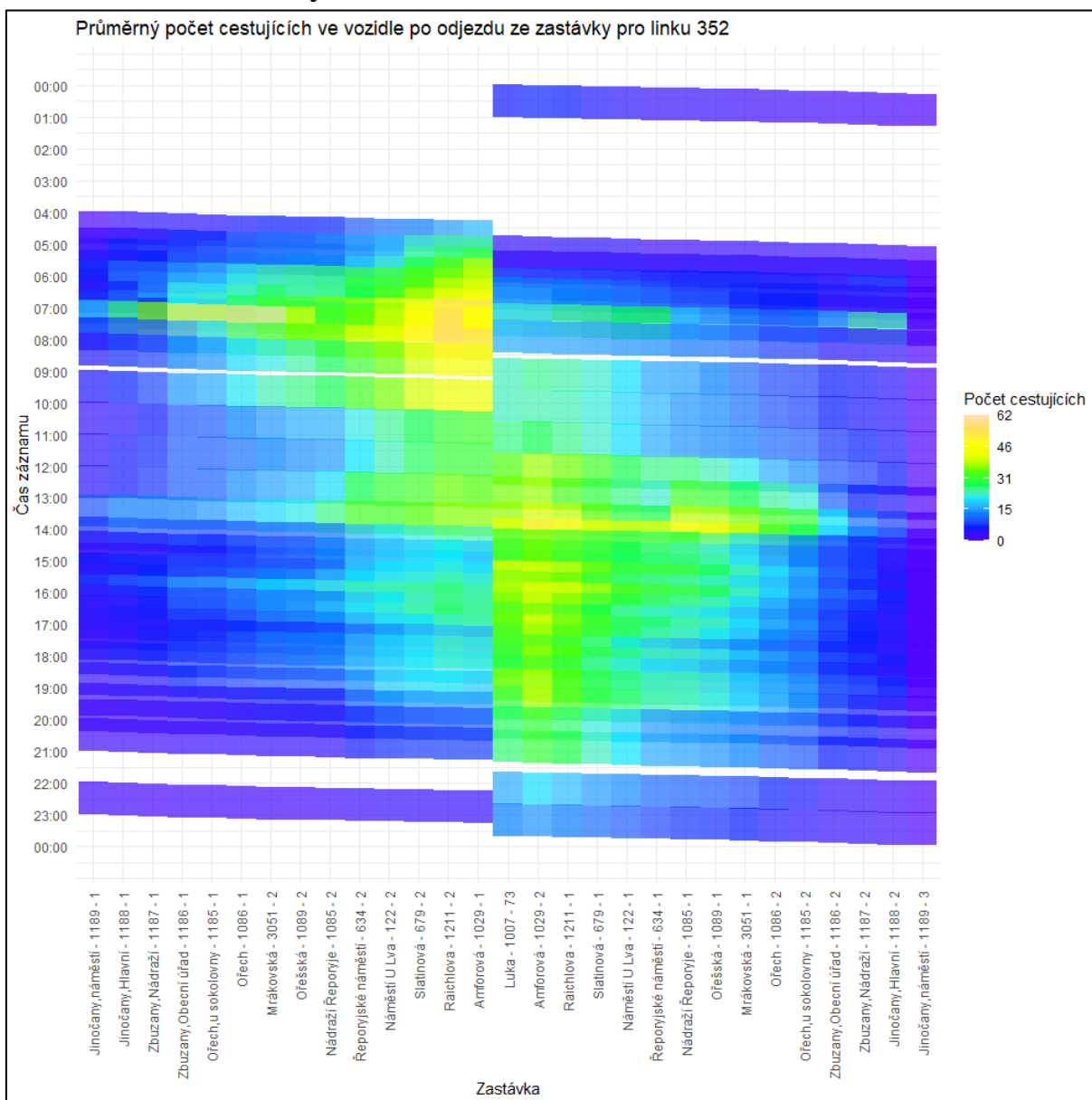
Na lince 352 (Praha, Luka –Jinočany, náměstí) byla převážně vypravována vozidla typu NB 12. V grafu níže lze pozorovat, že pravidelně byla spíše pokryta dřívější pořadí, trend však není výrazný.



Obrázek 47 - Zobrazení pokrytí a typů nasazených autobusů na lince 352

Celkové pokrytí linky systémem APC je 94,62 %. Počty cestujících zobrazované v následujících grafech jsou vynásobeny hodnotou $\frac{100}{94,62}$, aby co nejvíce odpovídaly skutečnému stavu.

4.5.1 Obsazenost linky 352 v čase



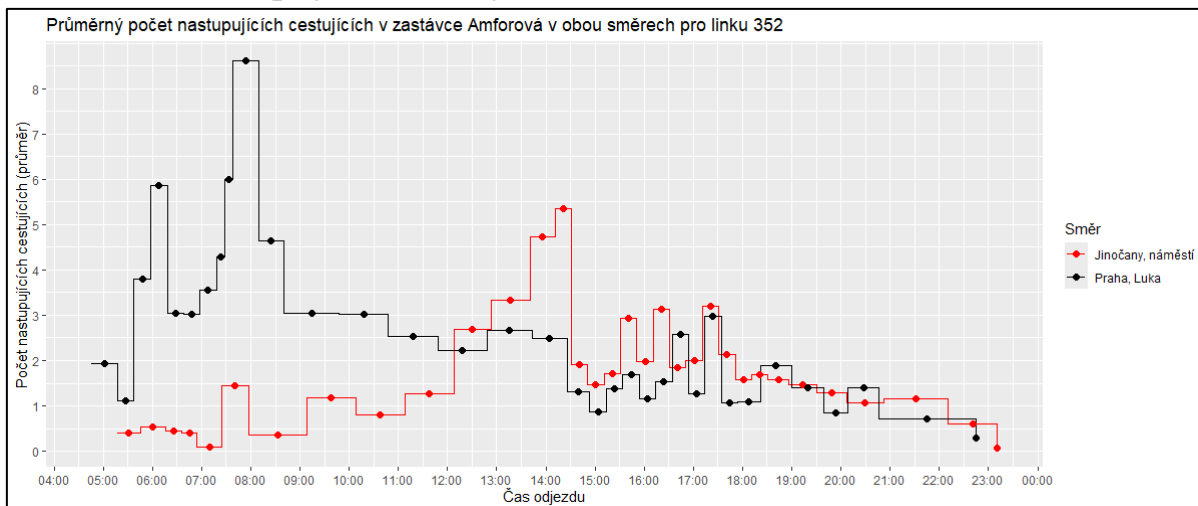
Obrázek 48 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 352

Linka 352 je příměstskou linkou obsluhující oblast Praha-jihozápad. Spojuje přílehlou obec Jinočany a metro Luka. V zastávce Luka (1007/73) se linka 352 takzvaně převleče a dál pokračuje jako linka 174 ve směru do zastávky Vypich. To je důvodem, proč konečnou zastávkou je zastávka Luka (1007/73) a následuje zastávka Amforová (1029/2), nikoli zastávka Luka v druhém směru (1007/71). Tento uzel náleží lince 174, která po zastavení v této zastávce dále pokračuje jako linka 352 s následující zastávkou Amforová (1029/2), tedy první zastávkou linky 352 ve směru do zastávky Jinočany, náměstí (1189/3). V grafu je sledována obsazenost čistě linky 352.

V grafu lze sledovat velmi výrazný průběh ranní špičky v časovém období 4:30-9:30 ve směru do zastávky Luka. Odpolední špička je znatelná v časovém rozmezí 13:00-19:00 ve směru do zastávky Jinočany, náměstí. Tento fakt potvrzuje domněnku, že linka 352 slouží cestujícím především jako doprava ráno do Prahy za prací a studiem a odpoledne doprava zpět do výchozího místa. Odpolední špička je méně ostrá než ranní, důvodem může být postupný čas konce vyučování ve školách.

Jelikož ve směru do zastávky Luka linka 352 dále pokračuje jako linka 174, nejsou po opuštění konečné zastávky Luka (1007/73) hodnoty počtu cestujících ve vozidle nulové.

4.5.2 Počet nastupujících cestujících v zastávce Amforová v čase



Obrázek 49 - Průměrný počet nastupujících cestujících v zastávce Amforová

V zastávce Amforová se průměrná hodnota počtu cestujících nastupujících do linky 352 pohybuje v rozmezí od 0 do 9. Tento graf výstupem odpovídá předešlému grafu. V ranní špičce jsou více vytíženy spoje ve směru do zastávky Luka, tedy směr na Prahu, odpoledne je naopak více vytížen směr z Prahy do zastávky Jinočany, náměstí.

5. Doporučení pro práci s APC daty

5.1 Vliv pokrytí na měření

Podstatnou částí této práce je analýza a zhodnocení stávajícího stavu pokrytí hlavního města Prahy systémem APC z různých úhlů pohledu. Základním předpokladem pro spolehlivost vyhodnocení dat jsou funkční a správně nastavená měřicí zařízení, která jsou pro sběr a uchování dat použita. Obecným předpokladem je, že čím vyšší míra pokrytí, tím více se lze na získaná data spolehnout. Míra pokrytí může být realizována a zlepšována na různých úrovních systému.

5.2 Doporučení při získávání APC dat

5.2.1 Pokrytí vozového parku

Jedním z výstupů práce je závěr, že vozový park je pokryt systémem APC z 47,22 %. Spodní hranice pokrytí je v návazné příloze Odbavovací a informační zařízení ve vozidlech PID stanovena na 33 %, tedy $\frac{1}{3}$ všech vozidel. Tato podmínka je splněna. Čím vyšší procento pokrytí vozového parku, tím více vozidel s možností měření je možno užívat v provozu a tím více dat je možno získat.

Z analýzy byla zjištěna informace, že minimálně některé typy nově pořizovaných vozidel jsou pořizovány včetně vybavení systémem APC ve 100% míře (nové vozy Crossway LE CITY 14.5M, nové vozy Urbino 10.5 IV), viz analýza souboru seznam_busu, Obrázek 20.

Pro maximální přesnost analýzy dat ze systému APC by bylo žádoucí, aby vozový park byl vybaven zařízeními APC ze 100 %.

5.2.2 Pokrytí linek

Dopravní podnik se zavazuje k rovnoměrnému využití vozidel se zařízením pro automatické počítání cestujících v rámci všech linek dopravce, viz kapitola 1.4 Obecné předpoklady v rámci PID (strana 9). Jak bylo zjištěno analýzou pokrytí jednotlivých linek v kapitole 3.4 Pokrytí jednotlivých linek v průběhu sledovaného období (strana 27), nebylo tomu tak.

Graf pokrytí jednotlivých linek (viz Příloha 1) ukazuje velké rozdíly v procentu pokrytí v závislosti na lince. Jedním z důvodů je fakt, že na některé linky se z určitých důvodů nasazují vozidla jednoho typu a může se stát, že tento typ autobusu nemá systémem APC vybaven ani jeden vůz. Příkladem může být linka 148 zajišťovaná vozidly typu Urbino 8.9 LE. Tento typ autobusu je svým provedením i kapacitou menší než standardní autobus délky 12 metrů, a specifickým potřebám provozu linky 148 plně vyhovuje. Z tohoto důvodu neexistuje jediný záznam o počtu cestujících využívajících linku 148 a pokrytí linky 148 je 0 %.

Naopak jsou i linky, na které byly nasazovány nově pořízené vozy, jejichž vybavení systémem APC je 100%. Příkladem takové linky je linka 101. Na ni byly nasazovány vozy typu Urbino 10.5 IV, viz kapitola 3. Analýza a vizualizace stávajícího pokrytí Prahy vozidly vybavenými APC (strana 16).

Poté existují linky, jejichž pokrytí systémem APC je skrz celé období nahodilé a existují značné rozdíly v rámci zkoumaných dnů. Příkladem takové linky je linka 230. Celkové procento pokrytí je 51,79 %. Na linku byla nasazována vozidla typu Urbino 8.9 LE, Urbino 10.5 IV a NB 12. Dle grafu bylo procento pokrytí některé dny nulové, jiné dny se pohybovalo okolo 30 % a jiné dny dosahovalo až ke 100 %.

Nakonec jsou v provozu i linky, jejichž pokrytí není 100 %, ale v rámci zkoumaného období byly hodnoty z jednotlivých dnů srovnatelné, například linka 109 (62,95 %). Dle fialové barvy v grafu skrze celé období lze konstatovat, že procento pokrytí bylo stálé.

V ideálním případě by zobrazený graf pokrytí měl celý tmavou barvu, což by znamenalo, že všechny provozované linky byly pokryty systémem APC ze 100 %. Nejen, že by nebyl rozdíl v procentu pokrytí napříč linkami, ale zároveň by toto procento již nebylo třeba zlepšovat.

Jelikož takového stavu nelze jednoduše a krátkodobě docílit, alespoň prozatímním cílem by mělo být, aby tento graf měl jednolitou barvu středně tmavého odstínu. Tedy nebylo by blízkým cílem, aby všechny linky byly pokryty ze 100 %, ale aby linky byly pokryty rovnoměrně a neexistovaly tak výrazné rozdíly mezi procentem pokrytí různých linek nebo různých dnů. V praxi by to znamenalo identifikovat linky, na které jsou pravidelně posílána vozidla, která nejsou systémem APC vybavena, a tato vozidla alespoň z části systémem vybavit a nebo je nahradit jinými vozidly, která již systém mají.

5.2.3 Pokrytí území hlavního města Prahy

Linky nebyly ve sledovaném období pokryty rovnoměrně, a to má za důsledek to, že nebyly rovnoměrně pokryty ani jednotlivé zastávky. Jak je možno vidět v mapě Vizualizace pokrytí jednotlivých zastávek s legendou (strana 20), nedá se říci, že by určitá část města byla systematicky více pokrytá než jiná, a to je dobře. Jednotlivé zastávky sice nejsou pokryty rovnoměrně, alespoň jsou však s ohledem na procento pokrytí rovnoměrně rozmístěny v rámci území.

Rovnoměrnost pokrytí území hlavního města Prahy úzce souvisí z předchozími dvěma body. Bude-li systémem APC rovnoměrně vybaven vozový park a budou-li vybavená vozidla rovnoměrně distribuována v rámci všech linek, bude pak rovnoměrně pokryto území města.

5.2.4 Pokrytí v rámci času

V kapitole 3.4.2 Detailní analýza pokrytí pořadí vybraných linek (strana 29) bylo zjištěno, že neexistuje striktně se opakující pravidelnost v pokrývání jednotlivých pořadí shodná pro všechny linky. Nelze tedy říct, že by například dopolední spoje byly pravidelně více pokryty než spoje odpolední, nebo naopak.

V takovém případě by bylo žádoucí navrhnout strategii pro identifikování časově nepokrytých úseků dne a přijít na způsob, jak je pokrýt. Vzhledem k tomu, že je však rozložení pokrytých pořadí linek nejednotné, doporučením může být snažit se, aby pravidelně méně pokrytá pořadí na různých linkách byla pokryta více i na úkor snížení pokrytí pořadí, která jsou nyní pravidelně pokryta ze 100 %.

5.3 Doporučení pro práci se získanými APC daty

5.3.1 Vypravení vozů s APC (apc-2023-02-06)

Tento CSV soubor dat o vypravení vozů vybavených systémem APC obsahuje 22 sloupců se specifickým významem. Každý sloupec má důvod, proč je daný údaj v něm o záznamu uchováván. Zorientování se ve významu sloupců proběhlo bez problémů.

Při analýze dat souboru bylo zjištěno několik anomálií, většina se při bližším zkoumání vysvětlila. Například chybějící údaje v záznamu ve sloupcích s názvy „prijezd“ a „zavreni“ (tyto chybějící údaje znamenají, že vozidlo zastávkou pouze projelo a nezastavilo, a nebo jde o konečnou zastávku) nebo nenulové počty cestujících ve vozidle v počáteční zastávce linky (to se děje, pokud linka plynule navazuje na jinou linku beze změny vozidla).

Jednou nesrovnalostí, kterou se nepodařilo objasnit, je problém převlékání linky 352 na linku 174 a zpět. V provozu to vypadá následovně: Do zastávky Luka, stanoviště 71, vjíždí linka 174, po opuštění zastávky Luka se ze spoje stane linka 352 a do následující zastávky linky 352 Amforová vjíždí již jako linka 352. V opačném směru se změna linky odehrává přesně opačným způsobem. Do zastávky Luka, stanoviště 73, vjíždí linka 352, po opuštění zastávky pokračuje dál ve směru do zastávky Vypich jako linka 174.

Tím pádem by se v zastávce Luka u linky 174 mělo objevit pouze stanoviště 71 a u linky 352 stanoviště 73. Data ukazují, že tomu tak není a existují záznamy obsahující údaje linka 174, zastávka Luka, stanoviště 73 i naopak linka 352, zastávka Luka, stanoviště 71. Nejspíš zde dochází k nestandardnímu chování sčítacího zařízení či jiné nesrovnalosti, jako například pochybení vlivem lidského faktoru. Bylo by dobré, kdyby byla provedena identifikace dalších zastávek, ve kterých probíhá převlečení jedné linky na jinou, a bylo by zjištěno, zda dochází k obdobnému jevu i tam.

Jinak se mi zdá soubor přehledný a přes jména sloupců dobře propojitelný s jinými soubory. Pro zvýšení přehlednosti a snadnější manipulaci s daty při porovnávání s jízdními řádů GTFS byl dodatečně vytvořen sloupec s názvem den a v něm dle data záznamu číslo dne v týdnu.

Do budoucna by bylo velmi příhodné, kdyby soubor též obsahoval alespoň základní údaje o všech spojích, nejen těch pokrytých APC. Tedy kdyby bylo zajištěno, že po vypsání dat ze zařízení APC budou tato data rovnou napárovaná s celkovým počtem provedených spojů. Odpadlo by tak složité párování dat ex-post a vyhledávání, jak to bylo (do minulosti), čímž může vzniknout množství dodatečných chyb.

5.3.2 Vybrané informace o vypravených spojkách (apc-vypraveni_dpp)

V souboru velmi často chybí uvedení evidenčního čísla vozidla ve sloupci v_vuz. Z toho důvodu je složité a někdy nemožné určit typ vozidla, který byl na dané pořadí nasazen. Chyba spočívá v neúplné databázi vypravení.

Rovněž ve sloupci v_vuz se objevilo 28 evidenčních čísel nasazených vozidel, která se nenacházejí v souboru seznam_busu, ve kterém by měla být sepsána veškerá vozidla provozovaná DPP. To je důvodem, proč se v grafu Rozdělení typů vozidel ve vozovém parku PID použitých v daném období (strana 26) zobrazuje 9,17 % Neznámo – u takových evidenčních čísel mi není dostupný záznam, o jaký typ vozu se jedná. Důvodem chybějících údajů bude pravděpodobně to, že soubor seznam_busu byl vyexportován na začátku ledna. Může se tedy stát, že na jaře byla v provozu vozidla, jejichž provoz byl do ledna ukončen, a tak seznam jejich evidenční čísla již neobsahuje.

Další nesrovnalostí je to, že soubor Vypravení vozů s APC obsahuje záznamy z linek 873 a 877, tyto linky se ale vůbec neobjevily v souboru apc-vypraveni_dpp. Ukázalo se, že linky sice byly v provozu, ale jedná se o neveřejné linky objednané soukromou institucí (soukromé školní linky).

5.3.3 Seznam jednotlivých typů vozů ve vozovém parku PID (seznam_busu)

Jak je zmíněno výše, soubor seznam_busu neobsahuje některá evidenční čísla, která se objevila v souboru apc-vypraveni_dpp. Otázkou je, který soubor je chybný nebo neúplný.

Dále bylo v souboru zjištěno, že se může stát, že k jednomu evidenčnímu číslu jsou přiřazeny dva typy vozidel. Tento problém se vyskytl u tří evidenčních čísel. Jedná se o tři muzejní vozidla výrobce Karosa, kterým byla po vyřazení z běžného provozu ponechána jejich evidenční čísla, která se nyní shodují s evidenčními čísly nových vozů [20]. To je důvodem výskytu vozů Karosa v grafu Rozdělení typů vozidel ve vozovém parku PID použitých v daném období (strana 26).

5.3.4 Jízdní řády v podobě GTFS

Pro práci s jízdními řády GTFS bylo žádoucí se nejprve seznámit s tím, jakým způsobem jsou sestavovány. Bylo třeba přesně stanovit, co je cílem z GTFS zjistit, a to počet spojů, které dle jízdního řádu projely všemi zastávkami poježděnými DPP. Zorientovat se, jaké údaje nese který textový soubor, které z nich jsou pro práci důležité a které nikoli.

Prvním zádrhelem byl fakt, že ačkoliv jízdní řády GTFS mají platnost dva týdny, velmi často se v nich některé údaje mění a musí být vydávány každý den. Proto je zvolen přístup porovnání tří analyzovaných dnů k jedné verzi jízdních řádů GTFS vydané v datum prvního z těchto tří dnů. Správně by každý jeden den měl být přiřazen k jízdnímu řádu GTFS, který byl vydán nejbližší tomuto datu.

Proces zpracovávání celkem 22 souborů GTFS, každého souboru zvlášť ve své vlastní databázi, byl zdoluhavý a ve výsledku velice náchylný pro vytvoření chyb z nepozornosti. Kdyby se analýza měla provádět pro delší časové období, důrazně bych doporučila najít jiný způsob zpracování. Přicházelo by v úvahu například dynamické načítání souborů nebo jiný způsob automatizovaného zpracování dat.

Dalším problémem se ukázalo být to, že jízdní řády GTFS jsou primárně vztaheny ke dnům v týdnu, nikoli ke konkrétním datům (ve smyslu slova datum). Tento fakt sám o sobě by nepředstavoval složitost, ba naopak by celý proces usnadňoval. Jenomže výjimky v provozu uvedené v souboru s názvem calendar_dates jsou vypsány pro konkrétní data. Jde například o mimořádné přidání spoje v den, kdy standardně nejede, nebo naopak vynechání spoje, který standardně jede. Nelze tedy jízdní řády GTFS porovnávat s údaji ze souboru Vypravení vozů s APC čistě na základě identifikování dne v týdnu.

Psaní dotazů v jazyku SQLite má z mého pohledu jednu velkou nevýhodu, a to takovou, že není možno v dotazu pracovat s proměnnými. Proto výpis počtu všech realizovaných spojů skrze zastávky byl jedním z posledních komplexních dotazů, které byly prováděny v databázi DB Browser. Od této chvíle bylo pokračováno v jazyce R v programu RStudio, který umožňuje proměnné používat.

Závěr

Tato práce je zaměřena na zhodnocení pokrytí hlavního města Prahy systémem automatického počítání cestujících (dále jen APC) ve vozidlech autobusů Dopravního podniku hl. m. Prahy (dále jen DPP). Sledovaným obdobím byl únor–červen 2023, a to výhradně dopravně inženýrsky významné dny: úterý, středa a čtvrtek. Pro analýzu byla uvažována data ze systému APC z celkem 65 dnů. Cílem práce je analyzovat specifické vlastnosti APC dat, analyzovat stávající pokrytí hl. m. Prahy vozidly vybavenými APC, analyzovat chování parametru obsazenosti linky v čase a formulovat doporučení pro práci s APC daty. Data pro analýzu byla poskytnuta společností ROPID.

Než je možno z dat vyvozovat jakékoliv závěry, je třeba ověřit jejich spolehlivost a určit, zda mají vypovídající hodnotu pro formulaci závěrů. V rámci práce byla data brána jako spolehlivá, ověření kvality proběhlo pouze na základě indikátoru kvality obsaženého v datech (sloupec vlajky). Ověření správnosti naměřených počtů cestujících a jiných údajů spadá pod kvalitu měřicího zařízení. Práce je tedy především zaměřena na posouzení rozložení zařízení APC v rámci hlavního města Prahy.

Pro získání informací o pokrytí hlavního města Prahy bylo třeba data ze systému APC dle dnů párovat na soubory jízdních řádů GTFS. Tak bylo možno mezi sebou porovnat pokryté a nepokryté spoje. Z jízdních řádů GTFS byl pro každou zastávku získán počet všech spojů, které projely danou zastávkou, počty odpovídají změnám jízdních řádů i mimořádným událostem v průběhu celého období. Toto porovnání se stalo zdrojem pro vizualizaci stávajícího pokrytí území (jednotlivých zastávek), analýzu pokrytí spojů v čase, vybavenost vozového parku a porovnání pokrytí jednotlivých linek.

Oběhů pokrytých systémem APC za sledované období vyjelo 31 567, nepokrytých 30 937. Celkové pokrytí oběhů je tedy 50,50 %. Vypravených vozů disponujících systémem APC bylo 561, vozů bez systému APC 628. Pokrytí užitého vozového parku je 47,22 %. Jelikož toto procento je nižší než procento pokrytí oběhů, lze zde vidět záměr posílat do provozu přednostně vozy vybavené systémem APC.

Dále ale bylo zjištěno, že vybavení systémem APC není rovnoměrně rozloženo v rámci typů autobusů. Existují linky, na které byl vypravován pouze specifický typ autobusu. Pokud tento typ nedisponuje systémem APC, má to za následek, že linka není pokryta. Příkladem je linka 148 a autobusy typu Urbino 8.9 LE. Bylo by vhodné, kdyby vozidla tohoto typu mohla být zpětně vybavena systémem APC, alespoň z postačujících 33 % - v případě potřeby i na úkor jiných typů autobusů, které jsou vybaveny z téměř 100 %, jako jsou autobusy typu Urbino 10.5 IV nebo BN 12. Z ekonomických důvodů je to však nemožné.

Autobusové zastávky nacházející se na území hlavního města Prahy obsluhované DPP nejsou pokryty rovnoměrně, alespoň jsou ale v rámci výše pokrytí v území rovnoměrně rozloženy. Přibližně polovina zastávek (1 167 zastávek) byla pokryta procentem vyšším než 50, druhá část (1 251 zastávek) byla pokryta procentem nižším než 50. Celkem 234 zastávkami za sledované období neprojel ani jeden spoj vybavený systémem APC a jejich pokrytí je 0 %. Naopak 49 zastávek bylo pokryto ze 100 %, tedy všechny spoje provozované DPP, které zastávkou za sledované období projely, byly systémem APC vybaveny.

V dokumentu Standardy kvality pro autobusy systému PID [7] je uvedeno, že podmínkou pro provoz vozidel se systémem APC je rovnoměrné využití těchto vozidel v rámci všech linek dopravce. Analýzou bylo zjištěno, že pokrytí linek rozhodně není rovnoměrné. V grafu pokrytí jednotlivých linek (viz Příloha 1) je na první pohled vidět, že pokrytí linek je velice rozdílné.

Celkem 13 pravidelných linek mělo celkové pokrytí nižší než 5 %, zatímco 11 pravidelných linek mělo pokrytí vyšší než 95 %. V ideálním případě by všechny linky měly pokrytí 100 %. Pokud ale nelze tohoto stavu jednoduše docílit, mělo by procento pokrytí linek být alespoň podobné. Je nepřípustné, aby mezi procentem pokrytí linek existovaly takto markantní rozdíly. Nejen že je to v rozporu s dokumentem Standardy kvality pro autobusy systému PID, zároveň by tím byla ovlivněna přesnost závěrů analýzy obsazenosti linek, kdyby měla být vytvořena pro každou linku. Takto musely být pro bližší analýzu vybrány specifické linky dosahující pokrytí alespoň 50 %.

Graf pokrytí jednotlivých linek vychází z dat za celé období a zobrazuje pokrytí oběhů linek v jednotlivé dny, lze tak porovnat i pokrytí linek v rámci času, nejen celkové pokrytí linky. Z grafu vyplývá, že pokrytí jedné linky v rámci času je obvykle pravidelně se opakující a nelze vidět snahu střídat vybavené vozy v rámci linek tak, aby se dostalo na všechny.

Sledování obsazenosti linky v čase bylo provedeno na čtyřech linkách odlišného charakteru s mírou pokrytí vyšší než 50 %. U každé linky byla vybrána jedna zastávka a byla provedena analýza počtu nastupujících cestujících v této zastávce do příslušné linky, rovněž vztaženo k času. Z důvodu pokrytí linek nižšího než 100 % byly počty dle míry pokrytí přenásobeny, aby hodnoty co možná nejvíce odpovídaly skutečnému stavu.

Jelikož zobrazované hodnoty obsazenosti linek jsou průměrnými hodnotami za celé období, z grafů Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro danou linku lze získat komplexní přehled o vytíženosti jednotlivých linek v čase. Všechny čtyři linky vykazují alespoň okrajově trend vyššího vytížení v období ranní a odpolední špičky. Dle povahy linky lze sledovat například tendenci cestujících dostávat se ráno směrem do centra Prahy a odpoledne naopak z centra (linka 133, Obrázek 42 a linka 352, Obrázek 48).

V závislosti na charakteru linky a místních cílech může být vidět vyšší míra obsazenosti i mimo standardní dopravní špičky (linka 119, Obrázek 39 a linka 201, Obrázek 45). Rovněž lze vidět nárůst či pokles obsazenosti v okolí přestupu na metro, záleží na směru cesty a umístění zastávky.

Vyšší obsazenost vozidla může vznikat před zastávkou skýtající možnost přestupu na metro, a naopak nízkou obsazenost vidíme v zastávkách následujících po přestupu – cestující přestupují z autobusu na metro (linka 201, zastávka Letňany, Obrázek 45). Stejně tak můžeme pozorovat nižší obsazenost vozidla před přestupem na metro a vyšší za přestupní zastávkou – cestující přestupují z metra na autobus (linka 201, zastávka Rajská zahrada, Obrázek 45).

Pro zlepšení kvality naměřených dat by bylo vhodné držet se následujících doporučení: Čím více vozů bude vybaveno systémem APC, tím více jich bude moct měřit v provozu. Není-li vozový park pokryt ze 100 % a není-li možno tohoto stavu jednoduše docílit, mělo by být rozložení vybavených vozidel v rámci denní doby rovnoměrné. Co se týká pokrytí jednotlivých linek, opět by mělo být cílem, aby jejich procento pokrytí bylo srovnatelné. Tímto způsobem lze docílit toho, že data získaná ze zařízení APC budou komplexní a dlouhodobým sběrem bude možno zajistit podklady pro analýzu o maximálně spolehlivé vypovídající hodnotě.

Zdroje

1. *abirail.cz* [online]. Případové studie, implementace automatického počítání cestujících do vozidel dopravního podniku, ©2024 [cit. 29. 4. 2024].
Dostupné z: https://www.abirail.cz/pripadove_studie/implementace-automatickeho-pocitani-cestujicich-do-vozidel-dopravniho-podniku/
2. *herman.cz* [online]. Jednotka počítání cestujících, ©2024 [cit. 29. 4. 2024].
Dostupné z: <https://www.herman.cz/cs/jednotka-pocitani-cestujicich-ucp-02>
3. *mikroelektronika.com* [online]. Zařízení pro počítání cestujících, ©2024 [cit. 29. 4. 2024].
Dostupné z: <https://www.mikroelektronika.com/cs/produkty/detail/zarizeni-pro-pocitani-cestujicich>
4. *vivotek.com* [online]. Zařízení, ©2024 [cit. 22. 5. 2024].
Dostupné z: <https://www.vivotek.com/sc9133-rtl>
5. *techtarget.com* [online]. Deep learning, Deep Neural Network, ©2024 [cit. 22. 5. 2024]. Dostupné z: <https://www.techtarget.com/searchenterpriseai/definition/deep-learning-deep-neural-network>
6. *vdv.de* [online]. VDV 457, ©2024 [cit. 29. 4. 2024].
Dostupné z: <https://www.vdv.de/457-v2.1-ses.pdf>
7. *pid.cz* [online]. Standardy kvality – autobusy, ©2024 [cit. 29. 4. 2024].
Dostupné z: https://pid.cz/wp-content/uploads/2018/04/Standardy_kvality_autobusy_2019.pdf
8. *pid.cz* [online]. Odbavovací a informační zařízení, ©2024 [cit. 29. 4. 2024].
Dostupné z: https://pid.cz/wpcontent/uploads/2022/07/Odbavovaci_a_informacni_zarizeni.pdf
9. *pid.cz* [online]. Seznam certifikovaných zařízení, ©2024 [cit. 29. 4. 2024].
Dostupné z: https://pid.cz/wp-content/uploads/2022/04/Seznam_certifikovanych_zarizeni.pdf
10. *iris-sensing.com* [online]. Produkt IRMA MATRIX, ©2024 [cit. 6. 5. 2024].
Dostupné z: <https://www.iris-sensing.com/products/irma-matrix/>
11. *iris-sensing.com* [online]. Technologie Time-of-flight, ©2024 [cit. 6. 5. 2024].
Dostupné z: <https://www.iris-sensing.com/products/time-of-flight-technology/>
12. *transitfeeds.com* [online]. Pražská integrovaná doprava, ©2024 [cit. 29. 4. 2024].
Dostupné z: <https://transitfeeds.com/p/prazska-integrovana-doprava/1106?p=19>
13. *developers.google.com* [online]. Transit, GTFS, ©2022 [cit. 16. 5. 2024].
Dostupné z: <https://developers.google.com/transit/gtfs>
14. *pid.cz* [online]. Open data, ©2024 [cit. 16. 5. 2024].
Dostupné z: <https://pid.cz/o-systemu/opendata/#h-gtfs>
15. *pid.cz* [online]. Sborník obsaditelnosti tramvají a autobusů, ©2016 [cit. 15. 7. 2024].
Dostupné z: https://pid.cz/wp-content/uploads/system/zkouska_obsaditelnosti_sbornik_2016.pdf
16. *sqlitebrowser.org* [online]. Download, ©2024 [cit. 23. 7. 2024]. Dostupné z: <https://sqlitebrowser.org/dl/>

17. *posit.co* [online]. Download, ©2024 [cit. 23. 7. 2024]. Dostupné z: <https://posit.co/download/rstudio-desktop/>
18. *cran.rstudio.com* [online]. Download and Install R, ©2024 [cit. 23. 7. 2024]. Dostupné z: <https://cran.rstudio.com/>
19. *seznam-autobusu.cz* [online]. Dopravní podnik hl. m. Prahy – autobusy, ©2024 [cit. 15. 7. 2024]. Dostupné z: <https://seznam-autobusu.cz/dopravce/dopravni-podnik-hl-m-prahy>
20. *seznam-autobusu.cz* [online]. Dopravní podnik hl. m. Prahy, 3709, ©2024 [cit. 23. 7. 2024]. Dostupné z: <https://seznam-autobusu.cz/seznam?operatorName=Dopravn%C3%AD+podnik+hl.+m.+Prahy&vc=3709&evn>
21. *tsk-praha.cz* [online]. Ročenka dopravy 2022, ©2024 [cit. 29. 4. 2024]. Dostupné z: <https://www.tsk-praha.cz/static/udi-rocenka-2022-cz.pdf>
22. *prg.aero* [online]. Letiště Václava Havla Praha, přehled letů, ©2024 [cit. 22. 7. 2024]. Dostupné z: <https://www.prg.aero/prehled-letu#/>

Seznam obrázků

Obrázek 1 - Umístění zařízení IRMA Matrix [10]	12
Obrázek 2 - Seznam souborů poskytnutých společnostmi ROPID	14
Obrázek 3 - Ukázka obsahu složky GTFS s platnostmi na 2 týdny	15
Obrázek 4 - Dotaz 3, dotaz pro počet APC spojů, které byly uskutečněny přes každou zastávku	18
Obrázek 5 - Dotaz 4, dotaz pro získání počtu všech spojů přes každou zastávku.....	19
Obrázek 6 - Dotaz 5, dotaz pro porovnání pokrytých a nepokrytých spojů.....	19
Obrázek 7 - Ukázka souboru POKRYTI_02_01	20
Obrázek 8 - Skript 1, skript pro vypočtení procenta pokrytí pro jednotlivé zastávky.....	20
Obrázek 9 - Ukázka souboru Procento pokrytí	21
Obrázek 10 - Vizualizace pokrytí jednotlivých zastávek s legendou, statický výřez interaktivní mapy	22
Obrázek 11 - Výřez mapy pokrytí, zastávka Bílá labuť, uzel 32/1	23
Obrázek 12 - Výřez mapy pokrytí, zastávka Náměstí Republiky, uzel 480/5	23
Obrázek 13 - Výřez mapy pokrytí, zastávka Park Rajská zahrada, 2789/1.....	24
Obrázek 14 - Vizualizace zastávek dle výše pokrytí (hranice 50 %) s legendou, statický výřez interaktivní mapy.....	25
Obrázek 15 - Výřez mapy pokrytí, centrum.....	26
Obrázek 16 - Dotaz 7, dotaz pro získání pořadí	26
Obrázek 17 - Dotaz 8, dotaz pro propojení souborů pořadí	26
Obrázek 18 – Dotaz 8, dotaz pro získání seznamu vypravených pořadí.....	27
Obrázek 19 - Dotaz 9, dotaz pro získání seznamu použitých vozidel	27
Obrázek 20 - Graf rozdělení typů vozidel ve vozovém parku PID použitých v daném období	28
Obrázek 21 - Pokrytí jednotlivých linek v průběhu sledovaného období s legendou, prvních 60 linek (celý graf viz Příloha 1).....	29
Obrázek 22 - Dotaz 10, dotaz pro získání tabulky obsahující informace o vypravených autobusech.....	31
Obrázek 23 - Ukázka tabulky obsahující informace o vypravených autobusech.....	32
Obrázek 24 - Ukázka záznamu s hodnotou CHYBA (v provozu více než jedno vozidlo)	33

Obrázek 25 - Ukázka záznamu, kdy byla nasazena dvě vozidla vybavena systémem APC, jedno z vozidel (apc_evč) se shoduje s plánovaným vozidlem (v_vuz).....	33
Obrázek 26 - Ukázka záznamu, kdy byla nasazena minimálně dvě vozidla, dvě s jistotou vybavená systémem APC, ani jedno z nich (apc_evč) se neshoduje s plánovaným vozidlem (v_vuz).....	33
Obrázek 27 - Zobrazení pokrytí a typů nasazených autobusů na lince 148	34
Obrázek 28 - Zobrazení pokrytí a typů nasazených autobusů na lince 113	35
Obrázek 29 - Záznam ze zdrojové tabulky pro zobrazení, 4. pořadí dne 1.2.2023	35
Obrázek 30 - Zobrazení pokrytí a typů nasazených autobusů na lince 137	36
Obrázek 31 - Zobrazení pokrytí a typů nasazených autobusů na lince 100	37
Obrázek 32 - Zobrazení pokrytí pořadí na lince 109.....	38
Obrázek 33 - Zobrazení pokrytí pořadí na lince 191	38
Obrázek 34 - Zobrazení pokrytí pořadí na lince 113	39
Obrázek 35 - Zobrazení pokrytí pořadí na lince 340.....	39
Obrázek 36 - Zobrazení pokrytí pořadí na lince 140.....	40
Obrázek 37 - Zobrazení pokrytí pořadí na lince 363.....	40
Obrázek 38 - Zobrazení pokrytí a typů nasazených autobusů na lince 119	43
Obrázek 39 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 119 ..	44
Obrázek 40 - Průměrný počet nastupujících cestujících v zastávce Navigátorů.....	45
Obrázek 41 - Zobrazení pokrytí a typů nasazených autobusů na lince 133	46
Obrázek 42 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 133 ..	47
Obrázek 43 - Průměrný počet nastupujících cestujících v zastávce Pernerova.....	48
Obrázek 44 - Zobrazení pokrytí a typů nasazených autobusů na lince 201	49
Obrázek 45 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 201 ..	50
Obrázek 46 - Průměrný počet nastupujících cestujících v zastávce Rajská zahrada.....	51
Obrázek 47 - Zobrazení pokrytí a typů nasazených autobusů na lince 352	52
Obrázek 48 - Průměrný počet cestujících ve vozidle po odjezdu ze zastávky pro linku 352 ..	53
Obrázek 49 - Průměrný počet nastupujících cestujících v zastávce Amforová	54

Seznam tabulek

Tabulka 1 - Ukázka výběru souborů GTFS pro jednotlivá data (datum)	16
Tabulka 2 - Spolehlivost všech obdržných záznamů.....	16
Tabulka 3 - Spolehlivost vybraných dnů, rozpis počtu záznamů s hodnotou vlajky jinou než 1	17
Tabulka 4 - Počty pokrytých a nepokrytých pořadí	27
Tabulka 5 - Pokrytí použitých vozidel autobusů	28
Tabulka 6 - Soupis blíže zkoumaných linek.....	41
Tabulka 7 - Soupis blíže zkoumaných zastávek	42

Seznam dotazů v jazyce SQLite

Dotaz 1

```
select * from apc_full where vlakky != 1;
```

Dotaz 2

```
select *,strftime('%w',datum) as "Den" from apc_full where den in ("2","3","4")
and linka not in ('873','877');
```

Dotaz 3

```
SELECT distinct nazev, zast, uzel, COUNT(*) AS count
FROM apc_final_ut_st_ct
WHERE datum = '2023-02-01'
GROUP BY uzel || '_' || zast
ORDER BY nazev, zast;
```

Dotaz 4

```
SELECT
    stops.stop_name,
    stops.asw_node_id,
    stops.asw_stop_id,
    COUNT(DISTINCT trips.trip_id) AS trip_count
FROM
    stop_times
JOIN
    trips ON stop_times.trip_id = trips.trip_id
JOIN
    stops ON stop_times.stop_id = stops.stop_id
JOIN
    calendar ON trips.service_id = calendar.service_id
JOIN
    routes ON trips.route_id = routes.route_id
LEFT JOIN
    calendar_dates ON (calendar_dates.service_id = calendar.service_id AND calendar_dates.date =
'20230201')
WHERE
    calendar.start_date <= '20230201'
    AND calendar.end_date >= '20230201'
    AND ((calendar.wednesday = true AND (calendar_dates.date IS NULL OR (calendar_dates.date != '20230201'
AND calendar_dates.exception_type = '2'))))
        OR (calendar_dates.date = '20230201' AND calendar_dates.exception_type = '1'))
    AND routes.route_type = 3
    AND trips.sub_agency_id = 1
    AND stops.location_type = 0
GROUP BY
    stops.stop_name,
    stops.asw_node_id,
    stops.asw_stop_id;
```

Dotaz 5

```
SELECT
    stop_name,
    GTFS_02_01_vystup.asw_node_id,
    GTFS_02_01_vystup.asw_stop_id,
    CAST(trip_count AS INTEGER) AS trip_count_numeric,
    CAST(APC_02_01_vystup.count AS INTEGER) AS apc_count_numeric,
    COALESCE(ROUND((CAST(APC_02_01_vystup.count AS INTEGER) /
NULLIF(CAST(GTFS_02_01_vystup.trip_count AS FLOAT), 0)) * 100, 2), 0) AS apc_percentage
FROM
    GTFS_02_01_vystup
LEFT JOIN
    APC_02_01_vystup ON GTFS_02_01_vystup.asw_node_id = APC_02_01_vystup.uzel
    AND GTFS_02_01_vystup.asw_stop_id = APC_02_01_vystup.zast;
```

Dotaz 6

```
select distinct datum ,strftime('%w',datum) as "Den" from apc_full where den in ("2","3","4");
```

Dotaz 7

```
select distinct datum, linkmen, poradi from apc_final_ut_st_ct
order by datum asc, linkmen asc, poradi asc;
```

Dotaz 8

```
SELECT v_datum, v_linka, v_poradi, km_linkmen, km_poradi
FROM vypraveni
LEFT JOIN linkmen_poradi_s_apc
ON vypraveni.v_datum = linkmen_poradi_s_apc.km_datum AND
    vypraveni.v_linka = linkmen_poradi_s_apc.km_linkmen AND
    vypraveni.v_poradi = linkmen_poradi_s_apc.km_poradi
WHERE v_linka > 99;
```

Dotaz 9

```

SELECT DISTINCT bus_actual, evč
FROM seznam_pouzitych_busu
LEFT JOIN apc_final_ut_st_ct
ON seznam_pouzitych_busu.bus_actual = apc_final_ut_st_ct.evč
WHERE apc_final_ut_st_ct.evč is not NULL
ORDER BY bus_actual ASC;

```

Dotaz 10

```

WITH bus_data AS (
    SELECT
        v_datum, v_linka, v_poradi, apc_linkmen, apc_poradi, v_vuz, apc_evč,
        CASE
            WHEN v_vuz IS NULL AND apc_evč IS NOT NULL THEN apc_evč
            WHEN apc_evč IS NULL AND v_vuz IS NOT NULL THEN v_vuz
            WHEN v_vuz IS NULL AND apc_evč IS NULL THEN NULL
            WHEN v_vuz = apc_evč THEN apc_evč
            ELSE 'CHYBA'
        END AS bus_actual
    FROM
        vypraveni
    LEFT JOIN
        linkmen_s_apc_s_evc ON vypraveni.v_datum = linkmen_s_apc_s_evc.apc_datum
        AND vypraveni.v_linka = linkmen_s_apc_s_evc.apc_linkmen
        AND vypraveni.v_poradi = linkmen_s_apc_s_evc.apc_poradi
    WHERE v_linka > 99
)
SELECT
    bus_data.v_datum, bus_data.v_linka, bus_data.v_poradi, bus_data.apc_linkmen,
    bus_data.apc_poradi, bus_data.v_vuz, bus_data.apc_evč, bus_data.bus_actual,
    CASE
        WHEN bus_actual = apc_evč THEN 'A'
        WHEN bus_actual = 'CHYBA' THEN 'A_CHYBA'
        ELSE 'N'
    END AS melo_apc,
    COALESCE (
        CASE
            WHEN bus_actual = 'CHYBA' THEN (
                SELECT seznam_busu.type
                FROM seznam_busu
                WHERE seznam_busu.number = bus_data.apc_evč
            )
            ELSE seznam_busu.type
        END, 'Neznámo') AS type
FROM
    bus_data
LEFT JOIN
    seznam_busu ON bus_data.bus_actual = seznam_busu.number
WHERE
    CASE
        WHEN bus_actual = apc_evč THEN 'A'
        WHEN bus_actual = 'CHYBA' THEN 'CHYBA'
        ELSE 'N'
    END IN ('A', 'N', 'CHYBA')
ORDER BY
    v_datum, v_linka, v_poradi;

```

Dotaz 11

```

select distinct bus_actual
from Pokryti_poradi
where v_linka = 148;

```

Seznam skriptů v jazyce R

Skript 1

```
library(dplyr)
library(tidyr)

stops <- read.csv("stops.csv")
POKRYTI_02_01 <- read.csv("POKRYTI_02_01.csv")
POKRYTI_02_02 <- read.csv("POKRYTI_02_02.csv")
...
POKRYTI_06_29 <- read.csv("POKRYTI_06_29.csv")

result_df <- POKRYTI_02_01 %>%
  inner_join(POKRYTI_02_02, by = c("stop_name", "asw_node_id", "asw_stop_id")) %>%
  inner_join(POKRYTI_02_07, by = c("stop_name", "asw_node_id", "asw_stop_id")) %>%
  ...
  inner_join(POKRYTI_06_29, by = c("stop_name", "asw_node_id", "asw_stop_id"))

result_df <- result_df %>%
  inner_join(stops, by = c("stop_name", "asw_node_id", "asw_stop_id")) %>%
  mutate(apc_vysledny = rowMeans(select(., apc_percentage), na.rm = TRUE),
         tripu_za_den_prumer = rowMeans(select(., trip_count_numeric), na.rm = TRUE),
         apc_za_den_prumer = rowMeans(select(., apc_count_numeric), na.rm = TRUE)) %>%
  replace_na(list(apc_za_den_prumer = 0)) %>%
  select(stop_name, asw_node_id, asw_stop_id, stop_lat, stop_lon, apc_vysledny, tripu_za_den_prumer,
         apc_za_den_prumer) %>%
  distinct() %>%
  arrange(stop_name, asw_node_id, asw_stop_id)

write.csv(result_df, "Procento_pokryti.csv", row.names = FALSE)
```

Skript 2

```
library(shiny)
library(leaflet)
library(dplyr)
library(RColorBrewer)

getColor1 <- function(values) {
  colors <- vector("list", length(values))
  for (i in seq_along(values)) {
    value <- values[i]
    if (value == 0) {
      colors[[i]] <- "grey"
    } else if (value == 100) {
      colors[[i]] <- "darkblue"
    } else {
      cols <- colorRampPalette(c("lightblue", "darkblue"))(101)
      colors[[i]] <- cols[value + 1]
    }
  }
  return(unlist(colors))
}

ui <- fluidPage(
  titlePanel("Porovnání spojů s APC a bez APC pro jednotlivé pražské zastávky"),
  mainPanel(
    leafletOutput("mojemapa"),
    br(),
    div(
      id = "legend",
      style = "position: fixed; top: 130px; right: 20px; z-index: 1000; background-color: #f2f2f2; padding: 10px;
border-radius: 5px; width: 290px;",
      h3("Legenda", style = "margin-top: -1px; color: black;"),
      div(
        id = "gradient",
        style = "background: linear-gradient(to right, lightblue, darkblue); height: 20px; position: relative;
margin-bottom: 10px;",
        tags$span("0%", style = "position: absolute; top: 20px; left: 0; font-size: 10px; width: 25%; text-align:
center; transform: translateX(-50%);"),
        tags$span("25%", style = "position: absolute; top: 20px; left: 25%; font-size: 10px; width: 25%; text-
align: center; transform: translateX(-50%);"),
        tags$span("50%", style = "position: absolute; top: 20px; left: 50%; font-size: 10px; width: 25%; text-
align: center; transform: translateX(-50%);"),
        tags$span("75%", style = "position: absolute; top: 20px; left: 75%; font-size: 10px; width: 25%; text-
align: center; transform: translateX(-50%);"),
        tags$span("100%", style = "position: absolute; top: 20px; left: 100%; font-size: 10px; width: 25%; text-
align: center; transform: translateX(-50%);")
      ),
      fluidRow(
        column(3, "APC = 0"),
        column(3, "APC = 0;100"),
        column(3, "APC = 100")
      ),
      fluidRow(
        column(3, "Počet:", textOutput("count_apc_0")),

```

```

        column(3, "Počet:", textOutput("count_apc_0_100")),
        column(3, "Počet:", textOutput("count_apc_100"))
      )
    )
  )
}

server <- function(input, output, session) {
  data <- read.csv("Procento_pokryti.csv", stringsAsFactors = FALSE)

  grouped_data <- data %>%
    group_by(stop_name, asw_node_id, asw_stop_id) %>%
    summarise(stop_lat = first(stop_lat),
              stop_lon = first(stop_lon),
              apc_vysledny = first(apc_vysledny),
              tripu_za_den_prumer = first(tripu_za_den_prumer),
              apc_za_den_prumer = first(apc_za_den_prumer))

  output$mojemapa <- renderLeaflet({
    leaflet(grouped_data) %>%
      addTiles() %>%
      addCircleMarkers(lng = ~stop_lon, lat = ~stop_lat,
                      radius = 5, fillOpacity = 0.7,
                      fillColor = ~getColor1(apc_vysledny),
                      color = NA,
                      popup = ~paste("Zastávka:", stop_name, "<br>",
                                     "Uzel:", asw_node_id, "/", asw_stop_id, "<br>",
                                     "Procento:", apc_vysledny, "% (", apc_za_den_prumer, ")", "<br>",
                                     "Průměr spojů za den:", tripu_za_den_prumer))

  })

  output$count_apc_0 <- renderText({
    sum(grouped_data$apc_vysledny == 0)
  })

  output$count_apc_0_100 <- renderText({
    sum(grouped_data$apc_vysledny > 0 & grouped_data$apc_vysledny < 100)
  })

  output$count_apc_100 <- renderText({
    sum(grouped_data$apc_vysledny == 100)
  })
}

shinyApp(ui = ui, server = server)

```

Skript 3

```

library(shiny)
library(leaflet)
library(dplyr)

getColor2 <- function(values) {
  colors <- vector("list", length(values))
  for (i in seq_along(values)) {
    value <- values[i]
    if (value < 50) {
      colors[[i]] <- "grey"
    } else {
      colors[[i]] <- "green"
    }
  }
  return(unlist(colors))
}

ui <- fluidPage(
  titlePanel("Rozdělení zastávek podle toho, zda procento pokrytí systémem APC dosáhlo 50 procent"),
  sidebarLayout(
    sidebarPanel(
      h3("Legenda"),
      fluidRow(
        column(6, "Procento < 50:", div(style = "background-color: grey; width: 30px; height: 20px;")),
        column(6, "Počet:", textOutput("count_below_50"))
      ),
      fluidRow(
        column(6, "Procento ≥ 50:", div(style = "background-color: green; width: 30px; height: 20px;")),
        column(6, "Počet:", textOutput("count_above_50"))
      )
    ),
    mainPanel(
      leafletOutput("mojemapa")
    )
  )
)

server <- function(input, output, session) {
  data <- read.csv("Procento_pokryti.csv", stringsAsFactors = FALSE)

```

```

grouped_data <- data %>%
  group_by(stop_name, asw_node_id, asw_stop_id) %>%
  summarise(stop_lat = first(stop_lat),
            stop_lon = first(stop_lon),
            apc_vysledny = first(apc_vysledny),
            tripu_za_den_prumer = first(tripu_za_den_prumer),
            apc_za_den_prumer = first(apc_za_den_prumer))

output$mojemapa <- renderLeaflet({
  leaflet(grouped_data) %>%
    addTiles() %>%
    addCircleMarkers(lng = ~stop_lon, lat = ~stop_lat,
                     radius = 5, fillOpacity = 0.7,
                     fillColor = ~getColor2(apc_vysledny),
                     color = NA,
                     popup = ~paste("Zastávka:", stop_name, "<br>",
                                   "Uzel:", asw_node_id, "/", asw_stop_id, "<br>",
                                   "Procento:", apc_vysledny, "% (", apc_za_den_prumer, ")", "<br>",
                                   "Průměr tripů za den:", tripu_za_den_prumer))
})

output$count_below_50 <- renderText({
  sum(grouped_data$apc_vysledny < 50)
})

output$count_above_50 <- renderText({
  sum(grouped_data$apc_vysledny >= 50)
})
}

shinyApp(ui = ui, server = server)

```

Skript 4

```

library(ggplot2)

data <- read.csv("Pokryti_poradi.csv")

unique_types <- c("Neznámo", "NB 18", "Urbino 10.5 IV", "NB 12", "BN 8.5", "BN 12", "Urbino 8.9 LE", "36BB
E'CITY", "B951E.1713", "Crossway LE CITY 13M", "B731.04", "Crossway LE 12.8M", "Crossway LE CITY 14.5M", "Crossway
LE CITY 12M", "NS 18", "Streetway 18M")
all_types <- factor(unique_types, levels = unique_types)
type_counts <- table(factor(data$type, levels = all_types))
total_buses <- length(data$type)
type_percentages <- round(100 * type_counts / total_buses, 2)
type_colors <- c("gray", "orange", "violet", "blue", "red", "yellow", "cyan", "magenta", "brown", "green",
"darkgreen", "purple", "turquoise", "coral", "maroon", "sienna")
data$type <- factor(data$type, levels = unique_types)

type_a_percentages <- sapply(unique_types, function(x) {
  type_data <- subset(data, type == x)
  if (nrow(type_data) == 0) return(NA)
  a_percentage <- sum(type_data$melo_apc == "A") / nrow(type_data) * 100
  return(round(a_percentage, 2))
})

ggplot(data, aes(x = "", fill = type)) +
  geom_bar(width = 1) +
  coord_polar("y", start = 0) +
  scale_fill_manual(values = type_colors, breaks = unique_types,
                    labels = paste(unique_types, " (", type_percentages, "%, s APC: ", type_a_percentages, "%)",
sep = "")) +
  theme_void() +
  theme(legend.position = "bottom") +
  labs(fill = "Typ vozidla", title = "Rozdělení typů vozidel")

```

Skript 5

```

library(ggplot2)
library(dplyr)

data$v_datum <- as.Date(data$v_datum)

coverage_data <- data %>%
  group_by(v_datum, v_linka) %>%
  summarise(
    total_trips = n(),
    apc_trips = sum(!is.na(km_linkmen))
  ) %>%
  mutate(coverage_percentage = (apc_trips / total_trips) * 100)

apc_coverage <- coverage_data %>%
  group_by(v_linka) %>%
  summarise(
    avg_coverage = mean(coverage_percentage, na.rm = TRUE)
  )

apc_coverage$label <- paste0(apc_coverage$v_linka, " ", sprintf("%.2f", apc_coverage$avg_coverage), "%")

```

```

available_dates <- unique(data$v_datum[data$v_datum %in% seq(min(data$v_datum), max(data$v_datum), by = "day")
& weekdays(data$v_datum) %in% c("Tuesday", "Wednesday", "Thursday")])

coverage_data$v_datum <- factor(coverage_data$v_datum, levels = available_dates)

ggplot(coverage_data, aes(x = factor(v_linka), y = v_datum, fill = coverage_percentage)) +
  geom_tile(color = ifelse(coverage_data$coverage_percentage < 12, "gray", "white"), size = 0.1) +
  scale_fill_gradient(low = "white", high = "blue", na.value = "gray", limits = c(0, 100)) +
  labs(x = "Linka", y = "Datum", fill = "Procento pokryti") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 5.5),
        axis.text = element_text(size = 5.5)) +
  scale_x_discrete(labels = apc_coverage$label) +
  scale_y_discrete(labels = format(available_dates, "%d. %m.)) +
  theme(axis.ticks.y = element_blank(),
        panel.grid.major.y = element_blank(),
        legend.position = "right")

```

Skript 6

```

library(ggnewscale)
library(ggplot2)

linka <- 100

pro_linku <- subset(data, v_linka == linka)
pro_linku$v_poradi_copy <- pro_linku$v_poradi
pro_linku$v_poradi_copy <- gsub("^\\d{1}[a-z]?$", "0\\1", pro_linku$v_poradi_copy)

data$v_datum <- as.Date(data$v_datum)
available_dates <- unique(data$v_datum[data$v_datum %in% seq(min(data$v_datum), max(data$v_datum), by = "day") &
weekdays(data$v_datum) %in% c("Tuesday", "Wednesday", "Thursday")])

unique_types <- c("Neznámo", "NB 18", "Urbino 10.5 IV", "NB 12", "BN 8.5", "BN 12", "Urbino 8.9 LE", "36BB
E'CITY", "B951E.1713", "Crossway LE CITY 13M", "B731.04", "Crossway LE 12.8M", "Crossway LE CITY 14.5M", "Crossway
LE CITY 12M", "NS 18", "Streetway 18M")
type_colors <- c("gray", "orange", "violet", "blue", "red", "lightblue", "cyan", "magenta", "brown", "green",
"darkgreen", "purple", "turquoise", "coral", "maroon", "sienna")
type_color_mapping <- data.frame(type = unique_types, color = type_colors[1:length(unique_types)])

pro_linku$appearance_order <- ave(seq(nrow(pro_linku)), pro_linku$v_datum, pro_linku$v_linka, pro_linku$v_poradi,
FUN = length)

available_dates <- rev(available_dates)

pro_linku$v_datum <- factor(pro_linku$v_datum, levels = available_dates)

ggplot(pro_linku, aes(x = as.factor(v_poradi_copy), y = v_datum)) +
  geom_tile(aes(fill = melo_apc, height = 0.9, width = ifelse(appearance_order > 1, 0.45, 0.9),
x = as.numeric(v_poradi_copy) - ifelse(appearance_order > 1, c(0.45/2, -0.45/2), 0)),
color = "black", position = "identity") +

  scale_fill_manual(name = "Vybavenost systémem APC",
values = c("lightgreen", "pink", "yellow"),
breaks = c("A", "N", "A_CHYBA"),
labels = c("S APC", "Bez APC", "S APC - ZMĚNA"),
na.value = "grey") +

  new_scale_fill() +
  geom_tile(data = subset(pro_linku, appearance_order == 1), aes(fill = type), width = 0.2, height = 0.7, color =
"black") +

  geom_tile(data = subset(pro_linku, appearance_order > 1), aes(fill = type, height = 0.7, width = 0.2,
x = as.numeric(v_poradi_copy) -
ifelse(appearance_order > 1, c(0.45/2, -0.45/2), 0)), color = "black", position = "identity") +

  scale_fill_manual(name = "Typ vozidla",
values = type_color_mapping$color,
breaks = type_color_mapping$type,
labels = type_color_mapping$type,
na.value = "grey") +

  labs(title = paste("Autobusová linka", linka), x = "Pořadí", y = "Datum") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 8),
        axis.text = element_text(size = 5.5),
        axis.ticks.y = element_blank(),
        panel.grid.major.y = element_blank(),
        legend.position = "right") +
  scale_y_discrete(labels = format(available_dates, "%d. %m.)) +
  scale_x_discrete(labels = function(x) gsub("^0", "", x))

```

Skript 7

```

library(ggplot2)
library(dplyr)

data <- read.csv("Pokryti_poradi.csv")

```

```

linka <- 100

pro_linku <- subset(data, v_linka == linka)

all_v_poradi <- data.frame(v_poradi = seq(min(pro_linku$v_poradi), max(pro_linku$v_poradi)))
pro_linku <- full_join(all_v_poradi, pro_linku, by = "v_poradi")
pro_linku$appearance_order <- ave(seq(nrow(pro_linku)), pro_linku$v_datum, pro_linku$v_linka, pro_linku$v_poradi,
FUN = length)

pro_linku$v_poradi_copy <- sprintf("%02s", pro_linku$v_poradi)

data$v_datum <- as.Date(data$v_datum)
available_dates <- unique(data$v_datum[data$v_datum %in% seq(min(data$v_datum), max(data$v_datum), by = "day") &
weekdays(data$v_datum) %in% c("Tuesday", "Wednesday", "Thursday")])

available_dates <- rev(available_dates)

pro_linku <- pro_linku %>%
  arrange(desc(v_datum))

ggplot(pro_linku, aes(x = as.factor(v_poradi_copy), y = factor(v_datum, levels = available_dates))) +
  geom_tile(aes(fill = melo_apc, height = 0.9,
    width = ifelse(appearance_order > 1, 0.45, 0.9),
    x = as.numeric(v_poradi_copy) - ifelse(appearance_order > 1, c(0.45/2, -0.45/2), 0)),
    color = "black", position = "identity") +
  scale_fill_manual(name = "Vybavenost systémem APC",
    values = c("lightgreen", "pink", "yellow"),
    breaks = c("A", "N", "A_CHYBA"),
    labels = c("S APC", "Bez APC", "S APC - ZMĚNA"),
    na.value = "grey") +
  labs(title = paste("Autobusová linka", linka), x = "Pořadí", y = "Datum") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 8),
    axis.text = element_text(size = 5.5),
    axis.ticks.y = element_blank(),
    panel.grid.major.y = element_blank(),
    legend.position = "right") +
  scale_y_discrete(labels = format(available_dates, "%d. %m.")) +
  scale_x_discrete(labels = function(x) gsub("^0", "", x)) +
  geom_text(aes(label = v_poradi), y = -0.5, size = 3, hjust = 0.5)

```

Skript 8

```

library(dplyr)
library(ggplot2)

data_full <- read.csv("apc_final_ut_st_ct.csv", header = TRUE)
data <- data_full
data <- subset(data, vlajky == 1)

pro_linku <- 174
pokryti <- 81.63
pro_uzel <- 872
pro_zast <- 8
pro_poradi <- 1

adjustment_factor <- 100 / pokryti

filtered_data <- data %>%
  filter(linka == pro_linku) %>%
  mutate(
    uzel_zast = paste(uzel, zast, sep = " - "),
    label = paste(nazev, uzel_zast, sep = " - "),
    adjusted_cestujicich = cestujicich * adjustment_factor,
    casjr = as.POSIXct(casjr, format = "%H:%M:%S", tz = "UTC"),
    casjr_numeric = as.numeric(casjr)
  )

initial_stop <- filtered_data %>%
  filter(datum == "2023-02-01", uzel == pro_uzel, zast == pro_zast)

if (nrow(initial_stop) > 1) {
  initial_stop <- initial_stop %>%
    slice(1)
}

initial_stop_time <- initial_stop$casjr

stop_order <- filtered_data %>%
  filter(datum == "2023-02-01", poradi == pro_poradi) %>%
  arrange(casjr) %>%
  distinct(uzel, zast, .keep_all = TRUE) %>%
  pull(label)

start_label <- paste(initial_stop$nazev, paste(initial_stop$uzel, initial_stop$zast, sep = " - "), sep = " - ")
stop_order <- c(start_label, stop_order[stop_order != start_label])

average_data <- filtered_data %>%

```

```

group_by(label, casjr_numeric) %>%
  summarize(average_passengers = mean(adjusted_cestujicich, na.rm = TRUE), .groups = "drop")

average_data <- average_data %>%
  mutate(label = factor(label, levels = stop_order))

average_data <- average_data %>%
  filter(!is.na(label))

hour_breaks <- seq(
  floor(min(average_data$casjr_numeric) / 3600) * 3600,
  ceiling(max(average_data$casjr_numeric) / 3600) * 3600,
  by = 3600
)

max_passengers <- max(average_data$average_passengers, na.rm = TRUE)
color_gradient <- topo.colors(10)

breaks <- seq(0, max_passengers, length.out = 5)

max_average_passengers <- average_data %>%
  filter(average_passengers == max(average_passengers, na.rm = TRUE))

print(paste("Bus stop with the most average passenger count:", max_average_passengers$label[1]))

ggplot(average_data, aes(x = label, y = casjr_numeric)) +
  geom_tile(aes(fill = average_passengers), width = 1, height = 3700, alpha = 0.7) +
  scale_fill_gradientn(
    colors = color_gradient,
    values = scales::rescale(breaks),
    breaks = breaks,
    labels = as.character(round(breaks)),
    name = "Počet cestujících"
  ) +
  scale_y_reverse(
    breaks = hour_breaks,
    labels = function(x) format(as.POSIXct(x, origin = "1970-01-01", tz = "UTC"), "%H:%M")
  ) +
  labs(x = 'Zastávka', y = 'Čas záznamu', title = paste('Průměrný počet cestujících ve vozidle po odjezdu ze
zastávky pro linku', pro_linku)) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))

```

Skript 9

```

library(dplyr)
library(ggplot2)

data_full <- read.csv("apc_final_ut_st_ct.csv", header = TRUE)
data <- data_full
data <- subset(data, vlajky == 1)

pro_linku <- 352
pokryti <- 94.62
pro_uzel <- 1029
pro_zast <- 1
smer <- "Praha, Luka"

pro_zast_2 <- 2
smer_2 <- "Jinočany, náměstí"

adjustment_factor <- 100 / pokryti

bus_stop_name <- data %>%
  filter(uzel == pro_uzel, zast == pro_zast) %>%
  select(nazev) %>%
  distinct() %>%
  pull()

filtered_data_nv <- data %>%
  filter(uzel == pro_uzel, zast == pro_zast, linka == pro_linku) %>%
  mutate(time_only = format(as.POSIXct(casjr, format = "%H:%M:%S"), "%H:%M:%S"))

average_nastup_nv <- filtered_data_nv %>%
  group_by(time_only) %>%
  summarize(average_nastup = mean(nastup, na.rm = TRUE) * adjustment_factor) %>%
  mutate(time_only = as.POSIXct(time_only, format = "%H:%M:%S"), direction = smer)

filtered_data_l <- data %>%
  filter(uzel == pro_uzel, zast == pro_zast_2, linka == pro_linku) %>%
  mutate(time_only = format(as.POSIXct(casjr, format = "%H:%M:%S"), "%H:%M:%S"))

average_nastup_l <- filtered_data_l %>%
  group_by(time_only) %>%
  summarize(average_nastup = mean(nastup, na.rm = TRUE) * adjustment_factor) %>%
  mutate(time_only = as.POSIXct(time_only, format = "%H:%M:%S"), direction = smer_2)

combined_data <- bind_rows(average_nastup_nv, average_nastup_l)

```

```
combined_data <- combined_data %>%
  arrange(direction, time_only) %>%
  group_by(direction) %>%
  mutate(mid_time = time_only + diff(c(time_only, time_only[length(time_only)])) / 2) %>%
  ungroup()

ggplot(combined_data, aes(x = time_only, y = average_nastup, color = direction)) +
  geom_step() +
  geom_point(aes(x = mid_time), size = 2) +
  scale_color_manual(values = setNames(c("black", "red"), c(smer, smer_2))) +
  labs(x = "Čas odjezdu", y = "Počet nastupujících cestujících (průměr)", color = "Směr") +
  scale_x_datetime(date_labels = "%H:%M", date_breaks = "1 hour") +
  scale_y_continuous(breaks = seq(0, max(combined_data$average_nastup, na.rm = TRUE), by = 1)) +
  ggtitle(paste("Průměrný počet nastupujících cestujících v zastávce", bus_stop_name, "v obou směrech pro linku",
    pro_linku))
```

Seznam příloh

Příloha 1 – Graf pokrytí jednotlivých linek

Příloha 1 - Graf pokrytí jednotlivých linek

