

Analýza dát zo cyklosčítačov v Bratislave

Tím: Jednotné programovacie družstvo a.s.

Marek Šugár
Lucia Ganajová
Tuan Dávid Nguyen Van
Jakub Novotný

Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
2026

Úvod

V posledných rokoch je z hľadiska urbánneho plánovania možné badať tendencie znižovania akejsi nadvlády áut v uliciach. Najmä v krajských mestách je zvyšovaný dôraz na budovanie infraštruktúr pre chodcov a cyklistov. Výnimkou nie je ani Bratislava, v ktorej je v posledných rokoch možné pozorovať nárast práve prepravy po dvoch kolesách. Pre úplnosť – po dvoch kolesách bez motora :)

Magistrát mesta aj z dôvodu ďalšej analýzy inštaloval od roku 2014 na niektorých úsekoch cyklotrás sčítače, ktoré s hodinovou frekvenciou akumulujú absolútne počty cyklistov, ktorí v danom časovom intervale prešli ktorýmkoľvek z dvoch smerov. Toto veľké množstvo dát vytvára priestor na hlbšie analýzy, ktoré bližšie dokážu opísať správanie cyklistov v Bratislave.

V tomto reporte je cieľom poukázať na zaujímavé interpretačné dôsledky vychádzajúce z dát. Dôraz je kladený na interpretáciu dát s rôznou frekvenciou a na spojitosti, resp. rozdielnosti medzi jednotlivými cyklotrasami. Postupne, v jednotlivých kapitolách reportu, je kladený dôraz na analýzu vplyvu vybraných kvantitatívnych vlastností počasia na vyťaženosť bratislavských cyklotrás. Bližšie sa preukazuje vyšší vplyv zmien počasia na špecifickú skupinu rekreačných, resp. provinčných, cyklotrás.

Špecifický dôraz je kladený na cyklotrasu na Vajanského nábreží, ktorej osadenie, ktorému predchádzalo zúženie doterajšej cestnej komunikácie o jeden pruh, vyvoláva aj dodnes výraznú spoločenskú diskusiu. Na základe veľkého množstva dát bolo možné preukázať viaceré hypotézy, ktoré do značnej miery odpovedajú na otázky opodstatnenia jej osadenia.

V neposlednom rade je dôraz kladený na analýzu krátkodobých, resp. dlhodobých, trendov vyťaženia cyklotrás. Je možné pozorovať viacero menej aj viac zjavných špecifik priebehu vyťaženia počas dní, týždňov až jednotlivých mesiacov roka.

Report vznikol ako jeden z podkladov riešenia semestrálneho projektu v rámci predmetu Princípy dátovej vedy v zimnom semestri 2025–2026 na Fakulte matematiky, fyziky a informatiky Univerzity Komenského v Bratislave.

v Bratislave, január 2026
autori

Obsah

Úvod	1
1 Štruktúra repozitára	3
2 Spojitosť vyťaženia cyklotrás a počasia Marek Šugár	4
2.1 Predspracovanie dát	4
2.2 Lineárna regresia	5
2.3 Interpretácia výsledkov	6
2.4 Penalizované lineárne regresie (LASSO, RIDGE)	8
3 Vyťaženosť cyklotrás v rámci dňa, týždňa, mesiaca, roku Lucia Ganajová	10
3.1 Dáta	10
3.2 Vyťaženosť počas dňa	10
3.2.1 Vyťaženosť jednotlivých trás	11
3.2.2 Porovnanie počas zimných a letných dní	12
3.3 Vyťaženosť počas jednotlivých mesiacov	12
3.4 Prediktívny model	13
3.4.1 Dáta	13
3.4.2 Použitý model - XGBoost	13
3.4.3 Výsledky	14
4 Klasifikácia cyklotrás na základe vyťaženia Jakub Novotný	16
4.1 Predspracovanie dát	16
4.2 Použité modely	16
4.2.1 Logistická regresia (základný model)	17
4.2.2 Random Forest Classifier	17
4.2.3 Neurónová sieť (LSTM)	18
4.3 Výsledky klasifikácie a vyhodnotenie	18
4.4 Analýza dôležitosti atribútov	20
5 Analýza vplyvu novej cyklotrasy na Vajanského nábreží Tuan Dávid Nguyen Van	22
5.1 Dáta	22
5.1.1 Čistenie a príprava dát	23
5.2 Regresný model	24
5.2.1 Model pre trasu Vajanského nábrežie	24
5.2.2 Porovnávacie modely pre trasy Páričkova a Viedenská	24
5.3 Výsledky	25
5.3.1 Porovnanie s dochádzkovými a rekreačnými trasami	25

Kapitola 1

Štruktúra repozitára

V prvom rade je vhodné ozrejmiť štruktúru repozitára, v ktorom sú dodané všetky zdrojové kódy a súbory, ktoré slúžili ako podklady k riešiu otázok v tomto reporte.

Každá samostatná kapitola, ktorá rieši určitú otázku, má svoj vlastný adresár v repozitári. Jednotlivé adresáre prináležia k kapitolám nasledovne:

- `/Weather Regression` – Kapitola 2: Spojitosť vyťaženia cyklotrás a počasia
- `/TemporalRoutePatterns` – Kapitola 3: Vyťaženosť cyklotrás v rámci dňa, týždňa, mesiaca, roku
- `/Routes Classifier` – Kapitola 4: Klasifikácia cyklotrás na základe vyťaženia
- `/Vajanského nábrežie` – Kapitola 5: Analýza vplyvu novej cyklotrasy na Vajanského nábreží

Okrem toho je súčasťou repozitára aj pomocný adresár `/Visuals of bicycle routes`, ktorý obsahuje vizualizácie všetkých analyzovaných cyklotrás adresár adresár a zároveň aj zdrojové kódy, ktoré pomocou knižnice `osmnx` slúžia ako zdroj vizuálov.

Kapitola 2

Spojitosť vyťaženia cyklotrás a počasia

Pre účely bližšieho porozumenia vplyvu stavu počasia v Bratislave na celkový počet absolútnych prejazdov po jednotlivých cyklotrasách je možné zvoliť viacero možných prístupov. V kapitole bol zvolený prístup tréningu vybraných algoritmov strojového učenia a následnej evaluácie predikcií modelov.

2.1 Predspracovanie dát

Dostupné dátové zdroje poskytujú s hodinovou frekvenciou informácie o počte prejazdov v oboch smeroch na danej cyklotrase. Dátové zdroje akumulujúce údaje o stave počasia boli avšak dostupné v spoľahlivej podobe iba s dennou frekvenciou. Z toho dôvodu bolo vhodné agregovať dáta z jednotlivých dostupných hodín prejazdov do konkrétnych dní. Pre účely priamočiarejšej a úspornejšej interpretácie má zmysel agregovať aj dáta z oboch smerov cyklotrasy dokopy – tým získavame počet denných prejazdov na cyklotrase celkovo. Pre účely interpretácie vplyvu počasia je to z apriórneho hľadiska postačujúce, z hľadiska dostupných dátových zdrojov jedna z mála možností.

Dátový zdroj zrkadlí počasie pomocou viacerých čiastkových premenných, niektoré viac, iné menej interpretačne uchopiteľné. Z relatívne bohatého dátového zdroja je možné vybrať reprezentatívnu podvzorku premenných – využívame prístup interpretačnej spojitosťi na základe istého poznania vplyvov počasia na tendenciu človeka *"ísť na bicykel"*.

Do našej využitej množiny premenných zaraďujeme údaje o teplote – **priemerná teplota**, **minimálna** a **maximálna teplota**. Na základe týchto dát je možné očakávať priamu úmernosť s absolútnym počtom prejazdov cyklotrasou. Vyššie teploty počas dňa pravdepodobne indukujú vyšší počet prejazdov, vyššiu tendenciu ľudí uprednostniť tento dopravný prostriedok pred napr. *autom*, *MHD* a pod.

Významný vplyv na vyťaženosť cyklotrasy, a na s ním prepojený sentiment obyvateľov voči cyklo-doprave, má určite **spád zrážok**. Do databázy premenných boli vložené obe dostupné metriky spádu zrážok – spád dažďových a snehových zrážok.

Okrem iného, významnými metrikami sú **rýchlosť vetra** a **tlak**. Prvá spomenutá metrika má dozaista vplyv na tendenciu ľudí uprednostniť dopravu bicyklom, spoločne s tlakom vzduchu, ktorý má významný vplyv na vyvíjanú biozáťaž a tendenciu ľudí vykonávať fyzickú záťaž na rámec, takpovediac, *nutnosti*.

Pri delení dát do zodpovedajúcich tréningových, validačných a testovacích množín dát je nutné zachovať chronologickú následnosť dát, keďže v podstate sa jedná o časové rady, a neuváženým prístupom k deleniu, je možné disponovať skreslenými, zväčša nadhodnotenými, výsledkami výkonnosti modelov.

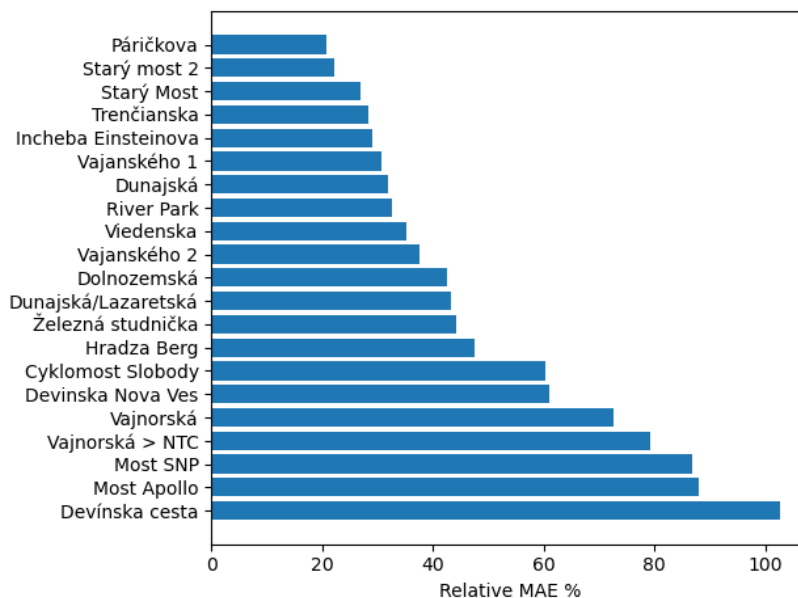
Z dôvodu variabilnej dĺžky monitorovaného časového intervalu pre rôzne cyklotrasy nie je možné pristupovať k deľbe pomocou fixných časových breakpointov. V projekte bol využitý prístup rozdelenia

celkového intervalu v pomere 6 : 3 : 1 pre tréningovú, validačnú a testovaciu množinu. Postupnosť dátových vstupov je zabezpečená zachovaním vzájomnej následnosti jednotlivých podintervalov. Tréningová množina predstavuje prvých 60% dát, validačná nasledujúcich 30% a zvyšok testovacia množina.

2.2 Lineárna regresia

Prvotnou ideou spomedzi klasických prístupov strojového učenia je lineárna regresia. Na tomto mieste je vhodné ozrejmiť aj metriku úspešnosti modelovania, na ktorú budeme reflektovať výkonnosti modelov. V kapitole je využitá **priemerná absolútna odchýlka**, normalizovaná zodpovedajúcim priemerom ako $\frac{MAE(Y; \hat{Y})}{\bar{Y}}$, kde Y označuje skutočnú hodnotu prejazdov, \hat{Y} predikcie modelu a \bar{Y} priemer skutočných hodnôt. Normalizácia je vhodná najmä z dôvodu relativizácie presnosti predikcie medzi rôznymi cyklotrasami. Metrika **priemernej absolútnej percentuálnej odchýlky (MAPE)** sa preukázala byť zväčša numericky nestabilná pri veľmi nízkych skutočných počtoch prejazdov, pričom pri relatívne presnej predikcii metrika indikovala abnormálne vysoké hodnoty chyby.

Výkonnosti modelovania pre jednotlivé cyklotrasy sú vizualizované na Obrázku 2.1.



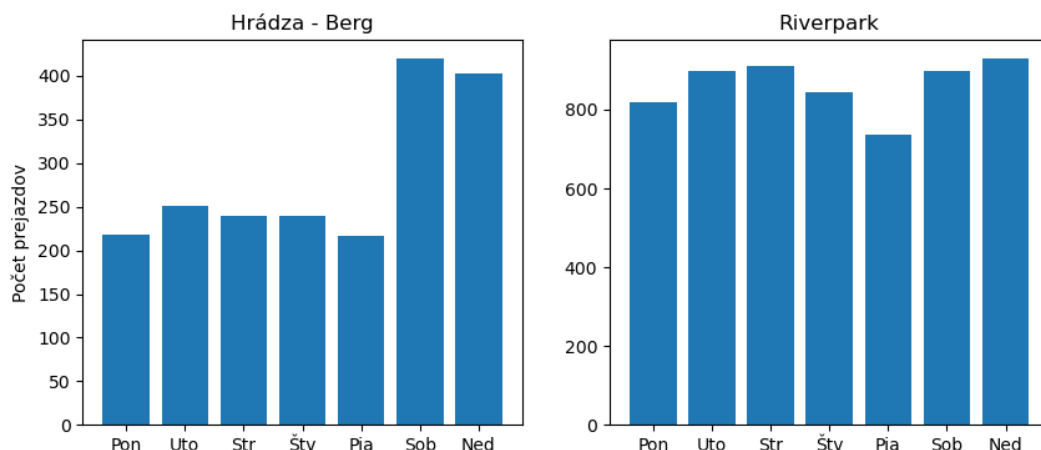
Obr. 2.1: Presnosť modelovania počtu prejazdov cyklotrasou lineárnou regresiou.

Je možné pozorovať pomerne široké spektrum presností - nižšie aj vyššie odchýlky. Odchýlky na úrovni približne 20% pozorujeme pri dvoch cyklocestách – **Páříčkova** a obe z trás na **Starom moste**. Nadpolovičná väčšina analyzovaných trás dosahuje relatívne odchýlky na úrovni pod 40%. Obzvlášť vyššie odchýlky naopak dosahuje cyklotrase spájajúca Karlovu Ves a Devín na **Devínskej ceste**, kde odchýlka presahuje hranicu 100%.

2.3 Interpretácia výsledkov

Na základe usporiadanej vizualizácie na Obrázku 2.1 je možné čiastočne načrtnúť odpoveď na otázku vplyvov počasia na absolútny počet prejazdov cyklotrasami. Toto vzostupné usporiadanie odchýlok preukazuje pravdepodobne klesajúcu mieru vplyvu počasia na početnosť prejazdov. Cyklotrasy s nižšou odchýlkou sú pravdepodobne silnejšie previazané s priebehom počasia a tie s vyššou odchýlkou naopak. Pozoruhodné je pozorovanie, že v prvej polovici cyklotrás s najvyššou mierou previazania sa nachádzajú len cyklotrasy v rámci centra, resp. širšieho centra alebo Petržalky. V druhej polovici, sa nachádzajú (s výnimkou Mosta SNP a Apollo) len cyklotrasy nachádzajúce sa skôr na perifériách mesta – napr. Cyklomost Slobody, Hrádza, Železná studnička a pod.

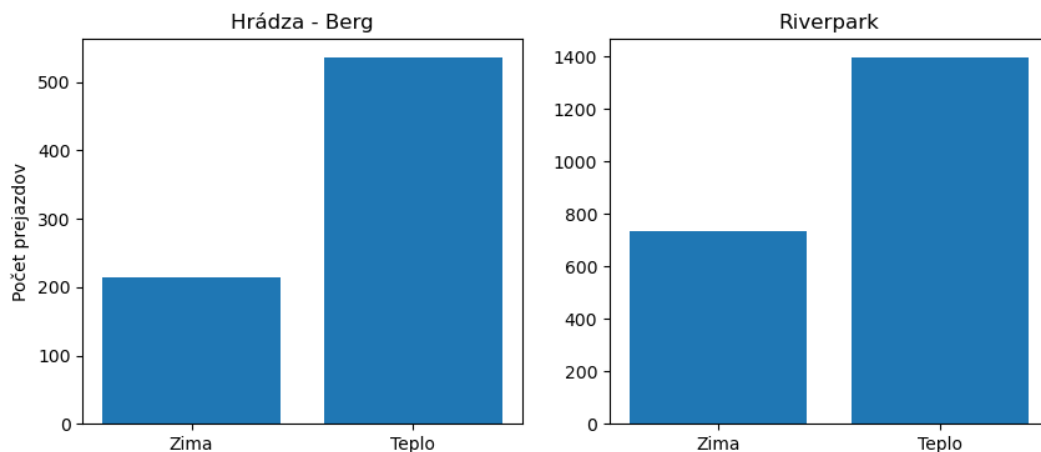
Táto skutočnosť je pomerne uchopiteľná. Cyklotrasy nachádzajúce sa v intenzívnejšie obývaných častiach mesta, pravdepodobne viac slúžia ako prostriedok dopravy do školy alebo zamestnania a teda môžu byť využívané takpovediac univerzálne – za každého počasia a približne v rovnakej miere počas týždňa. Opačne, periférnejšie cyklotrasy sú pravdepodobne využívané väčšinou rekreačne, pravdepodobne v rámci víkendov. Na Obrázku 2.2 je možné vidieť porovnanie dvoch cyklotrás, jednej v centre a druhej na periférii, z hľadiska priemerného počtu prejazdov počas týždňa, kde je zrejma koncentrickosť využívania cez víkend na periférnej trase a približne rovnomerné využívanie počas celého týždňa.



Obr. 2.2: Porovnanie priemerného počtu prejazdov na trasách počas celého týždňa.

Táto skutočnosť je aspoň čiastočným vysvetlením získaných výsledkov. Periférnejšie – rekreačnejšie trasy sú využívané skôr na voľnočasové jazdy, na čo má pravdepodobne vysoký vplyv počasia. Predpokladáme, že existuje veľmi nízke množstvo prejazdov na týchto trasách, napr. keď prší alebo sneží. Naopak, centrálnnejšie trasy sú využívané univerzálnejšie, ako prostriedok dopravy do zamestnania, resp. školy po celý rok – celý týždeň.

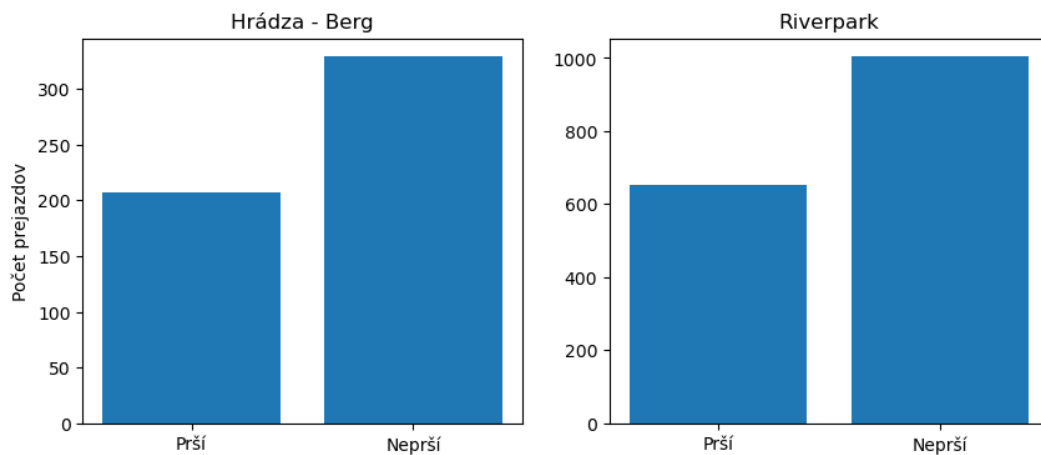
Táto skutočnosť je pomerne vhodne vizualizovaná na Obrázku 2.3, ktorý porovnáva priemernú využiteľnosť trás Riverpark a Hrádza - Berg v dňoch s teplotami $> 20^{\circ}\text{C}$ (Teplo) a $\leq 20^{\circ}\text{C}$ (Zima) počas dňa.



Obr. 2.3: Porovnanie priemerného počtu prejazdov na trasách na základe teploty.

Preukazuje sa, že pri analýze priemerného počtu prejazdov na základe teploty počas dňa pozorujeme v priemere pokles prejazdov na oboch trasách, na Hrádza - Berg o necelých 60% a na trase Riverpark o 47.44%. Toto správanie bolo preukázané vo všetkých trasách mimo centra – percentuálny pokles je vyšší než pri centrálnejších. Táto skutočnosť preukazuje, že trasy mimo centra sú pravdepodobne viac previazané na niektoré metriky počasia počas dňa.

Na druhú stranu je vhodné poukázať na odlišné správanie z tohto hľadiska pri inej metrike počasia. Na Obrázku 2.4 je vyobrazený priemerný počet prejazdov trasou v dňoch s nulovým a nenulovým spádom zrážok počas dňa.



Obr. 2.4: Porovnanie priemerného počtu prejazdov na trasách na základe spádu zrážok počas dňa.

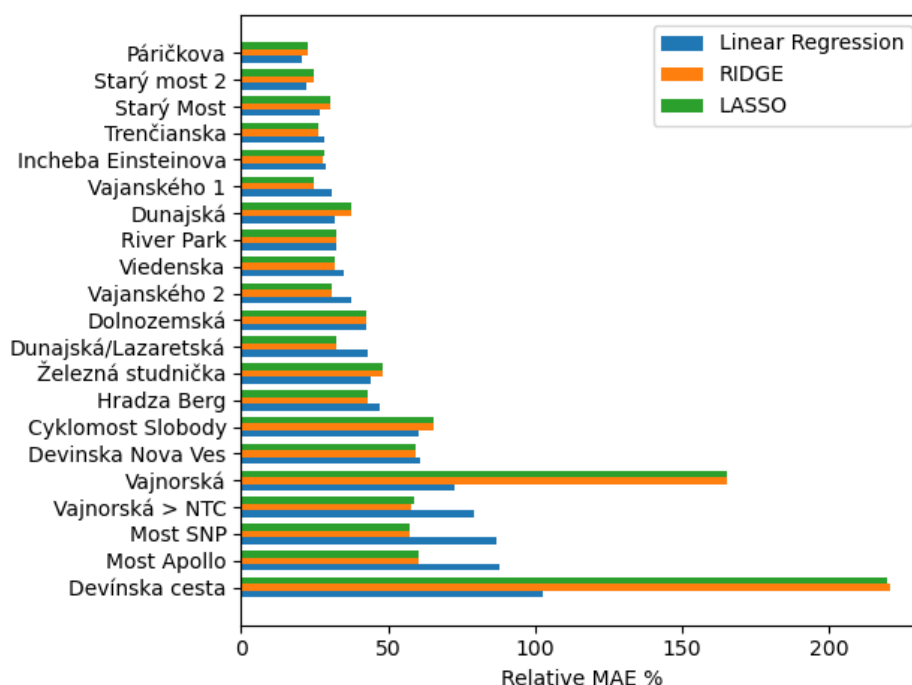
V tomto prípade nie je rozdiel v percentuálnych poklesoch ľahko badateľný. Na trase Hrádza - Berg pozorujeme pokles na úrovni 36.96% a na trase Riverpark 34.89%. Napriek rovnakému usporiadaniu, čiže vyšší pokles pri Hrádza – Berg, čo podporuje našu hypotézu, je rozdiel medzi jednotlivými trasami významne menší. To nás privádza k potenciálnemu použitiu modelovania s využitím istej formy výberu podvzorky premenných.

2.4 Penalizované lineárne regresie (LASSO, RIDGE)

V sekcii vyššie bolo preukázané, že vysoko pravdepodobne závisí, od toho, na základe ktorej metriky je diferencovaný vplyv počasia na počet prejazdov trasou. Z toho titulu môže byť vhodné využiť penalizovanú lineárnu regresiu a jej metriky úspešnosti modelovania porovnať s lineárnou regresiou. V tomto prípade má zmysel hypertunovať parameter modelu λ .

V tomto prípade sme zvolili prístup validácie modelov pre jednotlivé trasy na validačných dátach s rôznymi $\lambda \in \{0.01, 0.02, \dots, 2\}$ pre LASSO a $\lambda \in \{0.01, 0.02, \dots, 10\}$ pre RIDGE. Samotné horné a dolné hranice boli odvodené porovnávaním viacerých hodnôt odchýlok, analyzovaním intenzity regularizácie a nakoniec aj bližšou analýzou priebehu odchýlok s rastúcou λ -regularizáciou. Ako približne optimálna voľba λ bola zvolená následne hodnota s minimálnou relatívnou MAE.

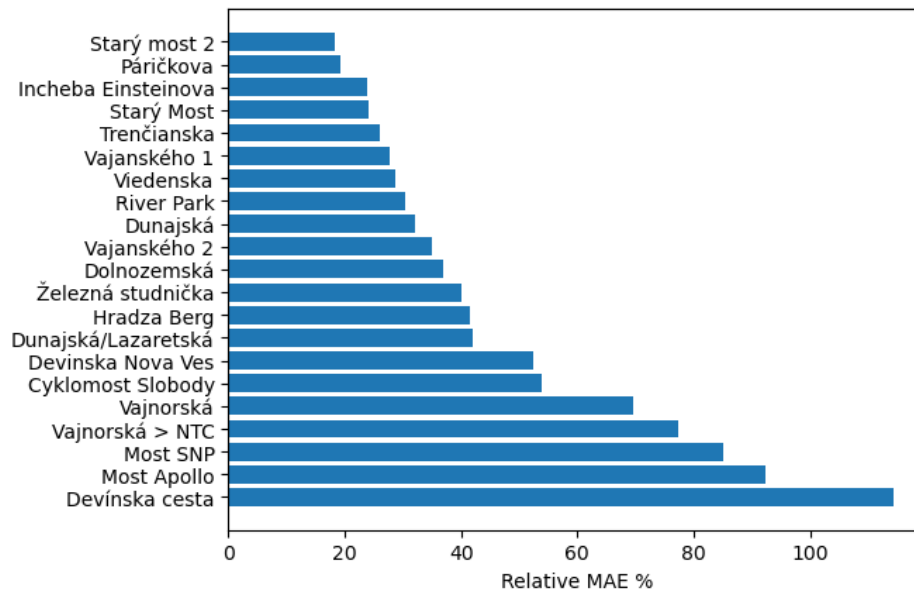
Na Obrázku 2.5 vyobrazujeme výslednú analýzu odchýlok pre jednotlivé regularizácie v porovnaní s lineárnou regresiou.



Obr. 2.5: Porovnanie výkonnosti predikcií počtu prejazdov na jednotlivých cyklocestách.

Je možné pozorovať približne rovnaké usporiadanie priemerných percentuálnych odchýlok pre jednotlivé cyklotrasy. Konkrétny trend nižšej priemernej odchýlky pre centrálnější trasy sa preukázal aj pri regularizačných algoritmoch. Pozoruhodné je však správanie pre trasy **Devínska cesta** a **Vajnorská**, kde oba typy regularizácie dosiahli obzvlášť vysokú odchýlku. Za spomenutie stojí aj to, že v drvivej väčšine trás je miera odchýlok vo všetkých troch algoritmoch veľmi podobná.

Prirodzenou voľbou oproti lineárnej regresii sa javí algoritmus založený na nelineárnej regresii. Pre účely tohto projektu bol v menšej miere využitý aj algoritmus **Random Forest Regressor**, avšak v relatívne obmedzenej forme – bez špecificky sofistikovaného tunovania všetkých jeho možných parametrov. Preukazuje sa, že úrovne odchýlok nie sú významne odlišné, resp. lepšie, než pri algoritmoch lineárnej regresie. Úrovne odchýlok pre jednotlivé cyklotrasy sú znázornené na Obrázku 2.6.



Obr. 2.6: Výkonnosť predikcii pomocou `RandomForestRegression` na jednotlivých cyklotrasách.

Je možné pozorovať relatívne obdobné usporiadanie úspešností regresie ako na Obrázku 2.5. Akési rozdelenie provinčných a centrálnych cyklotrás v jednotlivých poloviciach je taktiež pomerne zjavné.

Záverom je možné konštatovať viaceré závery. Prístupom tréningu nepenalizovanej, resp. penalizovanej, regresie sa preukázala istá prepojenosť počasia na cyklotrasy skôr v rámci centra mesta. Tu sa však preukázalo, že tento prístup nemusí byť nutne korektný z dôvodu absencie porovnateľne veľkého a rôznorodého množstva dát pre provinčné cyklotrasy mimo širšieho centra mesta, v ktorých sme na Obrázku 2.2 preukázali koncentrickosť cez víkendy. Tým sa ukázalo, že takéto porovnanie nie je nutne korektné, keďže samotná distribúcia dát v rámci jednotlivých dní nie je vyrovnaná.

Z toho dôvodu bol kladený dôraz na hlbšiu analýzu vplyvu počasia po jednotlivých metrikách počasia, ktoré môžu pravdepodobne mať vyšší vplyv na zmeny v správaní cyklistov. Preukázali sme, že vplyvom dažďových zrážok a teploty nad, resp. pod, hranicou 20°C . Je možné pozorovať vyšší pokles, resp. nárast priemerného počtu prejazdov trasou na trasách provinčných – napr. **Devínska cesta**, **Hrádza – Berg**, **Cyklomost Slobody** a i. Čím sme aspoň čiastočne načrtli alebo skôr odokryli viaceré ďalšie možné smery výskumu, ukazujúc, že trasy, nachádzajúce sa ďalej od mestskej zástavby, majú v istej miere vyššie previazanie na počasiu, ako tie v rámci centra mesta.

Kapitola 3

Časová vyťaženosť cyklotrás

V tejto kapitole sa zameriame na analýzu časovej vyťaženosť cyklotrás z rôznych časových perspektív – počas dňa, týždňa, mesiaca a roka. Cieľom je preskúmať, ako sa intenzita cyklistickej dopravy mení v priebehu týchto časových intervalov a či sú cyklotrasy využívané rovnomerne, alebo či medzi nimi existujú výrazné rozdiely.

3.1 Dáta

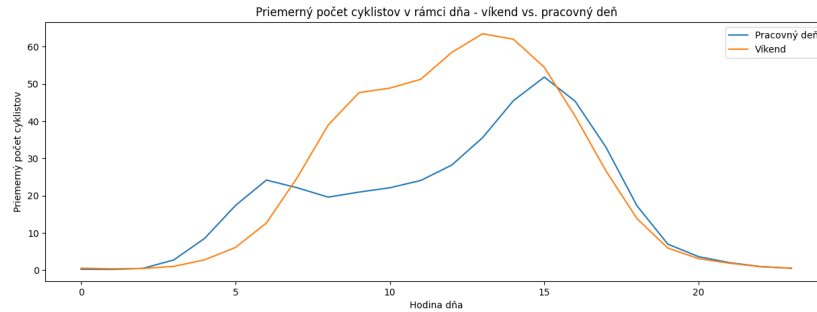
Pre analýzu časovej vyťaženosť sme pôvodné dáta rozšírili o viaceré časové atribúty, ktoré nám umožňujú podrobnejší pohľad na vyťaženosť cyklotrás. Medzi pridelenými atribútmi sú:

- `hour` – hodina dňa
- `dayofweek` – deň v týždni
- `dayofyear` – poradové číslo dňa v roku
- `month`, `quarter`, `year` – pre sledovanie dlhodobějších trendov
- `is_weekend` – binárny indikátor víkendu
- `is_winter`, `is_summer`, `is_spring`, `is_fall` - indikátory jednotlivých ročných období

Hodina dňa je kľúčová pre analýzu vyťaženosť v rámci jedného dňa, zatiaľ čo atribút `is_weekend` nám pomáha rozlíšiť vzory medzi pracovnými dňami a víkendmi, ktoré sú spravidla odlišné. Ďalšie časové ukazovatele nám umožňujú sledovať sezónne a ročné trendy.

3.2 Vyťaženosť počas dňa

Najmenšou sledovanou časovou mierkou je deň, preto začneme touto analýzou. Už na prvý pohľad je na Obrázku 3.1 vidieť výrazný rozdiel medzi vyťaženosťou počas pracovných dní a víkendov. Počas víkendov je na cyklotrasách zaznamenaný vyšší počet cyklistov ako počas týždňa. Ďalším rozdielom je, že počas pracovných dní majú počty cyklistov dva výrazné vrcholy: ráno okolo šiestej hodiny a popoludní okolo šiestnatej hodiny, čo naznačuje hlavne dochádzkovú dopravu do práce či školy. Naopak počas víkendov je počet cyklistov všeobecne vyšší a rozložený rovnomernejšie počas celého dňa, čo môže súvisieť s rekreačným využívaním cyklotrás.



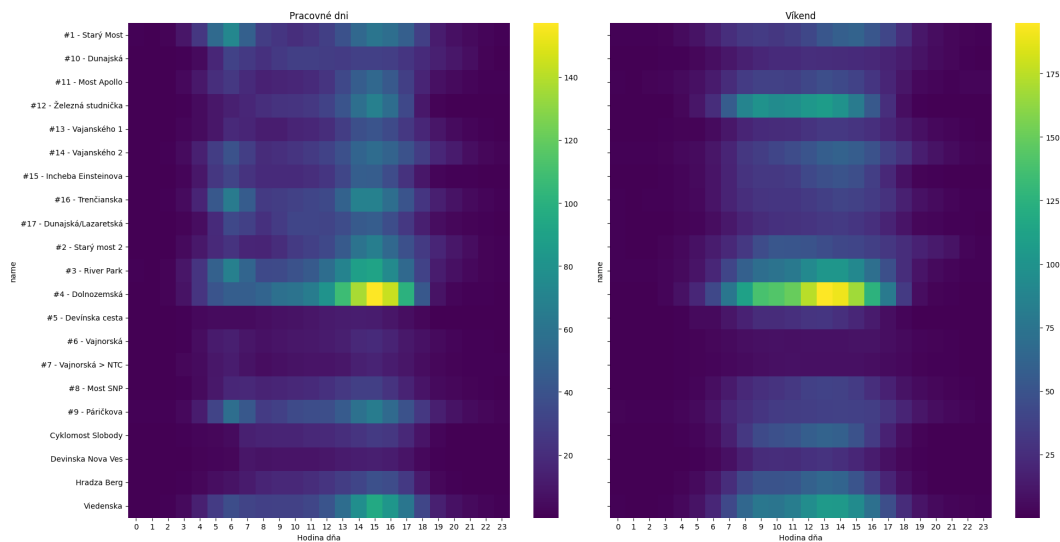
Obr. 3.1: Porovnanie priemernej vyťaženosti cyklotrás v rámci dňa počas pracovných dní a víkendu.

3.2.1 Vyťaženosť jednotlivých trás

Obrázok 3.2 rozlišuje vyťaženosť podľa jednotlivých cyklotrás počas pracovných dní a víkendov. Väčšina trás vykazuje rovnaký vzor – počas pracovných dní sú využívané najmä na dochádzku do práce či školy. Samozrejme existujú aj výnimky: Dunajská cyklotrasa alebo Cyklomost Slobody sa nevyznačujú dvoma výraznými vrcholmi počas pracovných dní, čo naznačuje skôr rekreačné využitie, nie dochádzkové. K týmto trasám patrí aj najvyťaženejšia zo všetkých - Dolnozemska cyklotrasa, ktorá je vyťažená počas celého týždňa, najmä v popoludňajších hodinách a taktiež počas víkendov, no to počas celého dňa.

Zaujímavé je tiež pozorovanie, že počas pracovných dní je popoludňajší vrchol (okolo šiestnastej hodiny) výraznejší než ranný. Hoci by sa mohlo zdať, že viac ľudí používa bicykel na cestu z práce, tento efekt pravdepodobne súvisí s vyšším počtom rekreačných cyklistov, ktorí sa po pracovnom dni venujú pohybu na bicykli.

Na základe dochádzkových vrcholov by sme jednotlivé trasy vedeli klasifikovať na primárne rekreačné a primárne dochádzkové.

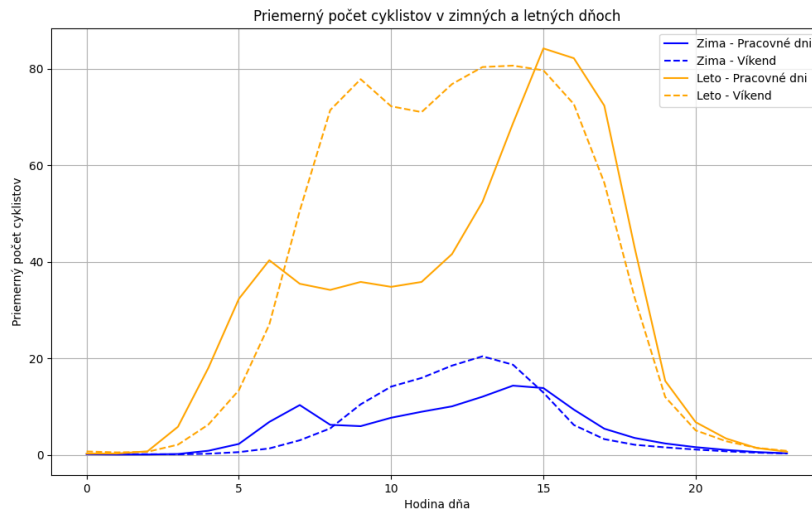


Obr. 3.2: Porovnanie priemernej vyťaženosti jednotlivých cyklotrás v rámci dňa počas pracovných dní a víkendu.

3.2.2 Porovnanie počas zimných a letných dní

Výrazné rozdiely možno pozorovať aj pri porovnaní zimnej a letnej sezóny. V zime je počet cyklistov výrazne nižší, čo súvisí s nepriaznivými poveternostnými podmienkami, nižšími teplotami a často aj s horším stavom ciest a chodníkov. Dáta ukazujú, že v zime sú 2 vrcholy vyťaženosti počas pracovného týždňa bližšie pri sebe, čo bude pravdepodobne dôsledkom kratších dní.

Počas víkendov sa počet cyklistov v zimnom období postupne zvyšuje s pribúdajúcimi hodinami dňa, pričom po približne trinástej hodine začína klesať. Po dvadsiatej hodine sa na cyklotrasách vyskytuje už len zanedbateľný počet cyklistov. V letnom období je naopak vyťaženosť vyššia počas celého dňa, s výraznejšími vrcholmi najmä v popoludňajších a večerných hodinách.

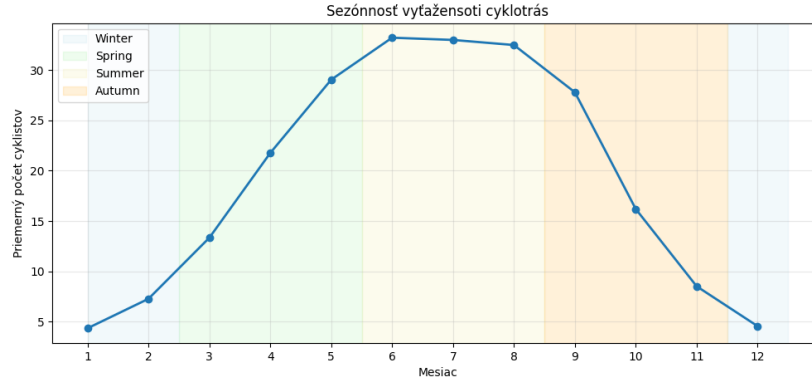


Obr. 3.3: Porovnanie priemernej vyťaženosti cyklotrás v rámci dňa počas pracovných dní a víkendu v zime aj v lete.

Pre jednotlivé trasy sme vypočítali aj tzv. **summer–winter ratio**, ktoré vyjadruje pomer priemerneho počtu cyklistov v letných a zimných mesiacoch a poskytuje tak kvantitatívnu mieru sezónnosti. Trasy Hrádza Berg, Devínska Nová Ves a Cyklomost Slobody sa ukázali ako lokality s najvýraznejším rozdielom medzi zimnou a letnou sezónou. Tento jav možno pripísať ich prevažne rekreačnému charakteru, v dôsledku čoho sú intenzívnejšie využívané najmä počas teplejších mesiacov.

3.3 Vyťaženosť počas jednotlivých mesiacov

Výsledky analýzy priemernej vyťaženosti cyklotrás v jednotlivých mesiacoch, ktoré môžeme bližšie vidieť na Obrázku 3.4 zodpovedajú očakávaným sezónnym trendom. Počet cyklistov sa zvyšuje v mesiacoch s vyššími dennými teplotami a naopak klesá počas zimného obdobia. Najvyššiu vyťaženosť dosahujú cyklotrasy v letných mesiacoch, pričom maximum je zaznamenané najmä v mesiaci jún. Hodnoty vyťaženosti počas letnej sezóny sú však pomerne rovnomerne rozložené. Naopak, najnižšia vyťaženosť je pozorovaná v januári.



Obr. 3.4: Porovnanie priemernej vyťaženosťi cyklotrás počas štyroch ročných období.

Na porovnanie sezónnosti jednotlivých cyklotrás sme vypočítali sezónnu variabilitu vyjadrenú **koefficientom variability**

$$CV = \frac{\text{smerodajná odchýlka}}{\text{priemer}}.$$

Poradie cyklotrás podľa tejto metriky sa výrazne nelíši od výsledkov získaných pomocou *summer-winter ratio*. Najvyššie hodnoty sezónnej variability opäť dosahujú rovnaké tri cyklotrasy, čo potvrdzuje konzistentnosť oboch prístupov pri hodnotení sezónnych rozdielov.

3.4 Prediktívny model

3.4.1 Dáta

Použité dáta sa oproti analýze časovej vyťaženosťi výrazne neodlišujú. Jedinou zmenou je úprava reprezentácie atribútu `dayofweek`, ktorý bol pôvodne vyjadrený ako celé číslo na škále 0–6.

Pri metódach strojového učenia však takáto reprezentácia cyklických dát vytvára umelú diskontinuitu medzi nedeľou a pondelkom, hoci tieto dni v týždni na seba bezprostredne nadväzujú. Z tohto dôvodu sme atribút `dayofweek` nahradili cyklickou reprezentáciou dňa v týždni pomocou trigonometrického enkódovania. Konkrétne boli vytvorené dva nové atribúty: `dayofweek_sin` a `dayofweek_cos`, ktoré umožňujú zachytiť cyklický charakter dní v týždni.

Na rozdiel od experimentu v Kapitole 2 sme sa rozhodli zahrnúť aj dáta z obdobia pred a počas pandémie ochorenia COVID-19. Keďže ide o úlohu modelovania časových radov, tri roky dát (od roku 2022) nie sú postačujúce na spoľahlivé natrénovanie a vyhodnotenie modelu. Keďže aj napriek tomu sú niektoré cyklotrasy otvorené až od neskorších rokov, môžeme očakávať aj veľmi nepresné predikcie, spôsobené malým množstvom tréningových dát, na ktorých sa model nestihol dostatočne naučiť všetky vlastnosti a zachytiť sezonalitu dát.

3.4.2 Použitý model - XGBoost

Na modelovanie vyťaženosťi cyklotrás bol zvolený model XGBoost (Extreme Gradient Boosting), ktorý patrí medzi ensemble metódy založené na rozhodovacích stromoch. XGBoost využíva princíp gradientného boostingu, pri ktorom sa model skladá z postupne tréňovaných stromov, pričom každý ďalší strom sa snaží korigovať chyby predchádzajúcich modelov.

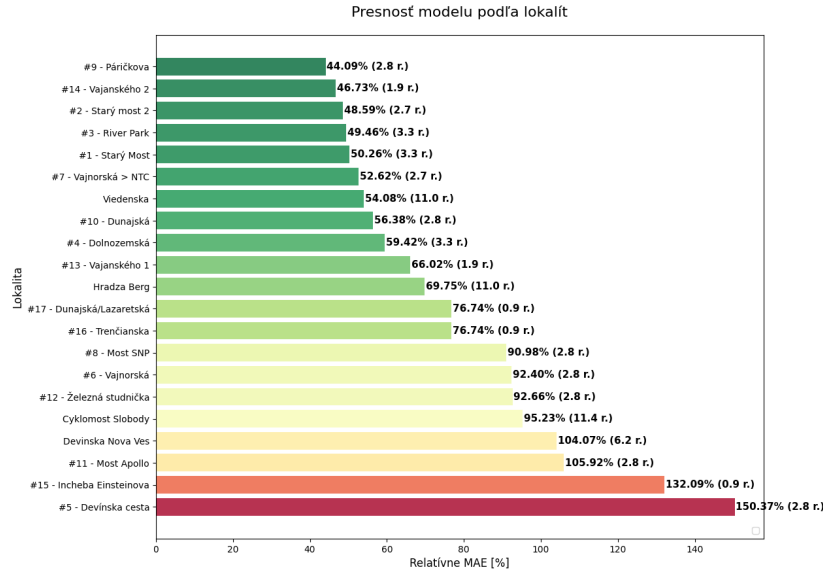
Rozhodovacie stromy sú síce flexibilné a dobre interpretovateľné, no často trpia problémom preučenia. XGBoost tento problém zmierňuje zavedením regularizácie, ako aj mechanizmov ako *early*

stopping, ktoré umožňujú zastaviť tréning v momente, keď sa výkonnosť modelu na validačných dátach prestane zlepšovať.

Vďaka týmto vlastnostiam je XGBoost vhodný pre modelovanie komplexných nelineárnych vzťahov v časových a sezónnych dátach, akými sú aj dáta o vyťaženosti cyklotrás.

3.4.3 Výsledky

Výsledky modelu sú znázornené na Obrázku 3.5, kde je pre každú cyklotrasu uvedená hodnota relatívneho MAE. Zároveň je v grafe zobrazené aj časové rozpätie dát, z ktorých bol model pre jednotlivé trasy trénovaný a validovaný.



Obr. 3.5: Výsledky modelu XGBoost na testovacej množine.

Z výsledkov môžeme pozorovať niekoľko zaujímavých skutočností.

Po prvé, intuitívne by sme očakávali, že cyklotrasy s kratším časovým radom budú dosahovať vyššie hodnoty relatívneho MAE z dôvodu menšieho množstva tréningových dát. Tento predpoklad sa však nepotvrďuje vo všetkých prípadoch. Pri detailnejšom pohľade na časovú vyťaženosť jednotlivých trás zistíme, že cyklotrasa *Párikčikova*, ktorá dosahuje najlepšie skóre, je síce trénovaná na relatívne malom objeme dát, avšak jej vyťaženosť má veľmi pravidelný a stabilný denný profil. Naopak, trasa *Devínska cesta*, ktorá obsahuje dáta z porovnateľného časového obdobia, vykazuje najhoršie výsledky. Tento jav je spôsobený výraznou nepravidelnosťou dát, vysokými výkyvmi v počte cyklistov a obdobiami s nulovou alebo takmer nulovou vyťaženosťou.

Po druhé, jPo druhé, pozornosť si vyžaduje použitá hodnotiacia metrika – relatívne MAE. Táto metrika vyjadruje veľkosť chyby modelu v percentách vzhľadom na priemernú hodnotu cieľovej premennej. V niektorých prípadoch však môže pôsobiť zavádzajúco. Ak je priemerný počet cyklistov na trase veľmi nízky alebo sa blíži k nule, aj malá absolútna chyba môže viesť k extrémne vysokým hodnotám relatívneho MAE.

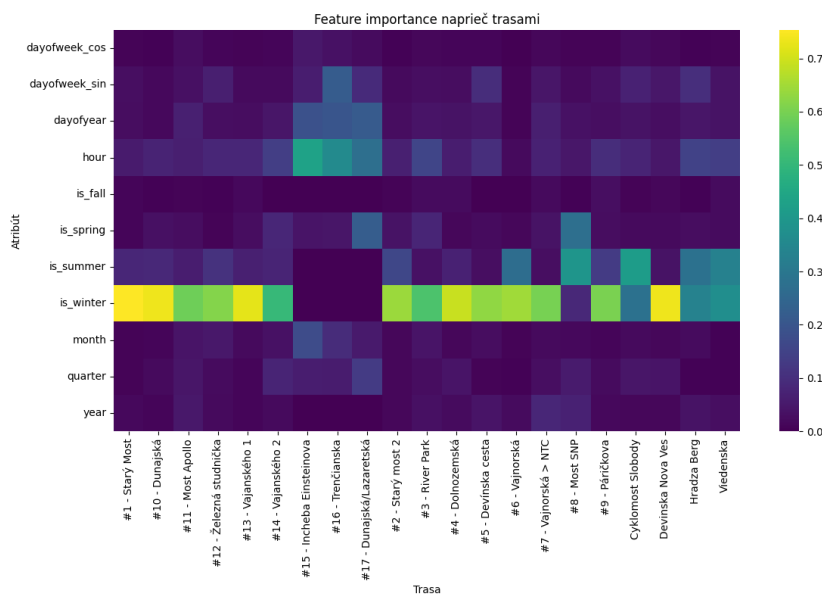
V analyzovaných dátach sme však nezaznamenali výraznú koreláciu medzi vysokými hodnotami relatívneho MAE a nízkym priemerným počtom cyklistov. Vo väčšine prípadov sú vysoké hodnoty tejto metriky spôsobené najmä nepravidelnosťou časových radov a slabou opakovateľnosťou vzorcov v dátach.

Dôležitosť atribútov

V tejto časti sme analyzovali atribúty, ktoré model XGBoost považoval pri predikcii za kľúčové. Na obrázku 3.6 je znázornená normalizovaná dôležitosť jednotlivých príznakov pre každú trasu, čo umožňuje ich vzájomné porovnanie.

Z výsledkov je zrejmé, že tri lokality sa výrazne odlišujú od ostatných: Incheba Einsteinova, Trenčianska a Dunajská/Lazaretská. V týchto prípadoch nemá atribút `is_winter` takmer žiadnu váhu. Pri bližšej analýze vstupných dát zisťujeme, že tento jav je spôsobený dĺžkou dostupných časových radov – pre tieto trasy disponujeme údajmi až z roku 2025. Keďže tréningová a validačná množina pre tieto senzory nepokrývala zimné mesiace v dostatočnej miere, model nedokázal zachytiť sezónny trend spojený so zimou. V dôsledku toho sa algoritmus pri predikcii spolieha primárne na atribút `hour` (hodina dňa), ktorý vykazuje stabilnú dennú periodicitu nezávislú od ročného obdobia.

Naopak, pri trasách s dlhšou históriou, ako sú Cyklomost Slobody či Viedenská, vidíme vysokú dôležitosť atribútu `is_winter` v kombinácii s `is_summer` a `hour`. To potvrdzuje predpoklad, že cyklistická doprava v Bratislave je silne sezónna a v zime očakávame výrazný pokles cyklistov. Zaujímavým zistením je aj nízka dôležitosť atribútov `year` a `quarter`, čo naznačuje, že medziročná dynamika nárastu počtu cyklistov je menej významná.



Obr. 3.6: Dôležitosť atribútov pri predikcii vyťaženia cyklotrás.

Kapitola 4

Klasifikácia cyklotrás na základe vyťaženia

Cieľom tejto kapitoly je zistiť, či môžeme určiť konkrétny názov cyklotrasy len na základe toho, ako sa využíva. Každé miesto v Bratislave má svoje vlastnosti – niektoré sú hlavne na dochádzanie do práce (sú tu veľké návaly ráno a poobede počas pracovných dní), iné sú viac na voľný čas (viac ľudí cez víkendy a poobede). V tejto časti použijeme viacero metód strojového učenia, konkrétne logistickú regresiu, náhodné lesy a neurónovú sieť. Úspešná klasifikácia by ukázala, že jednotlivé cyklotrasy majú jedinečné a predvídateľné ‘správanie’.

4.1 Predspracovanie dát

Na klasifikáciu sme surové dáta upravili na formát, ktorý môžu použiť všetky použité metódy, vrátane neurónovej siete. Tento proces mal tieto kroky:

- **Agregácia a čistenie:** Dáta sme prepočítali na hodinové intervaly a chýbajúce hodnoty nahradili nulou. Odstránili sme všetky týždne, kde neboli kompletne záznamy, aby sme predišli skresleniu.
- **Vytváranie sekvencií:** Dáta sme rozdelili na týždenné úseky (168 hodín). Vstup pre jednu vzorku obsahuje kombináciu oboch smerov jazdy („Z“ a „Do“), čo dáva $168 + 168$ hodnôt pre každú klasifikovanú vzorku.
- **Normalizácia:** Všetky vstupné hodnoty sme štandardizovali pre *logistickú regresiu* a škálovali na interval $[0, 1]$ pre *neurónovú sieť*.
- **Encoding:** Názov cyklotrasy sme najprv očíslovali pomocou *LabelEncoder* a potom previedli na binárne vektory pomocou *to_categorical* z knižnice *keras*.

4.2 Použité modely

V tejto časti sme testovali, ako dobre rôzne algoritmy dokážu klasifikovať cyklotrasy. Začali sme jednoduchými modelmi, ktoré spracovávali dáta po jednotlivých záznamoch, a prešli sme k zložitým neurónovým sieťam, ktoré spracovávajú časové rady ako celok.

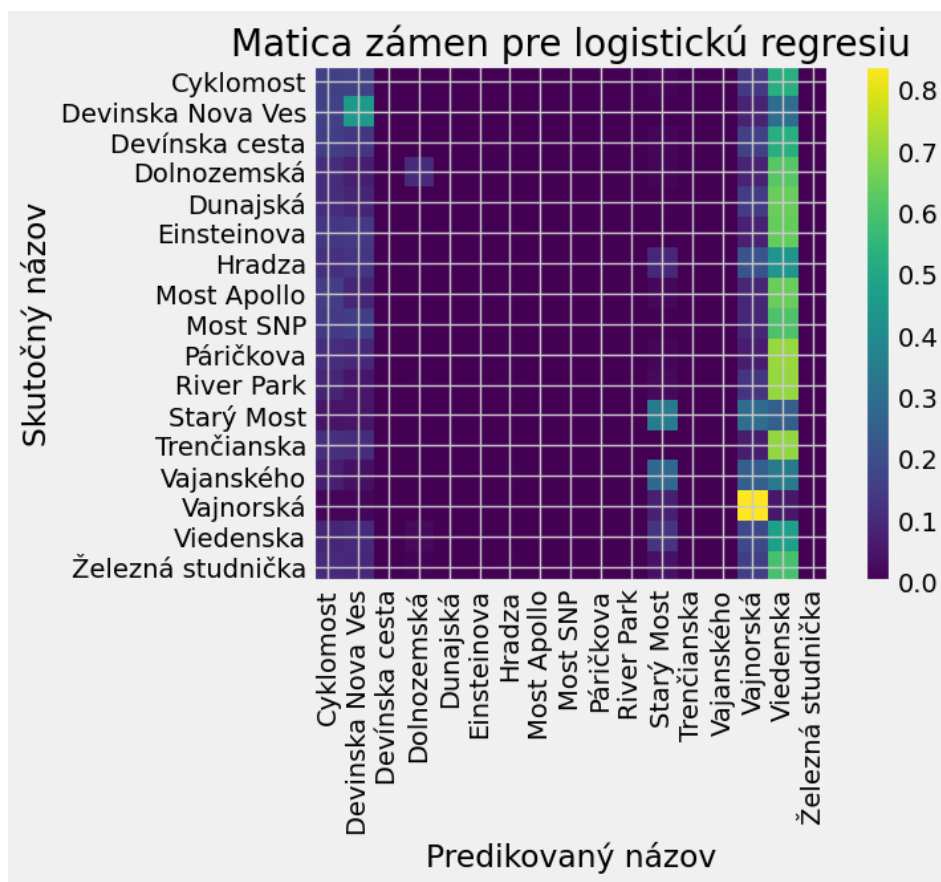
Na rozdiel od predchádzajúcich kapitol sme tentokrát nezohľadnili počasie, aby sme to zjednodušili. Chceli sme zistiť, či môžeme trasu určiť len na základe vnútorných vzorcov vyťaženia (hodina, deň, počet cyklistov).

Modely, najmä tie jednoduchšie, sa snažili dosiahnuť úspech tým, že predikovali najčastejšie sa vyskytujúce trasy, čo skresľovalo presnosť. Model sa ‘naučil’, že najlepšie je tipovať trasy s najväčším počtom záznamov, preto sme pristúpili k využívaniu **váhovania tried**.

4.2.1 Logistická regresia (základný model)

Ako základ sme použili logistickú regresiu. Tento model sme trénovali na pôvodnej štruktúre dát, kde každý riadok bol samostatný záznam. Vstupné údaje: ['POCET_Z', 'POCET_DO', 'hodina', 'den_v_tyzdni', 'mesiac']. Časové údaje sme vzali z dátumu a času.

Model dosiahol nízku presnosť (19%). To značí, že posudzovať jednotlivé hodiny zvlášť nestačí, pretože mnohé trasy majú v určitých hodinách podobný počet cyklistov a ťažko sa dá vyčítať nejaké trendy. Zároveň je na Obrázku 4.1 možné pozorovať, že model sa skôr naučil početnosti trás, pretože predikoval najčastejšie trasy v dátach, dokonca sa vôbec nesnažil predikovať tie menej vyskytujúce. To nám ukázalo problém nevyváženosti dát, v matici nie je takmer nič na diagonále - tzn. veľmi nízku presnosť.

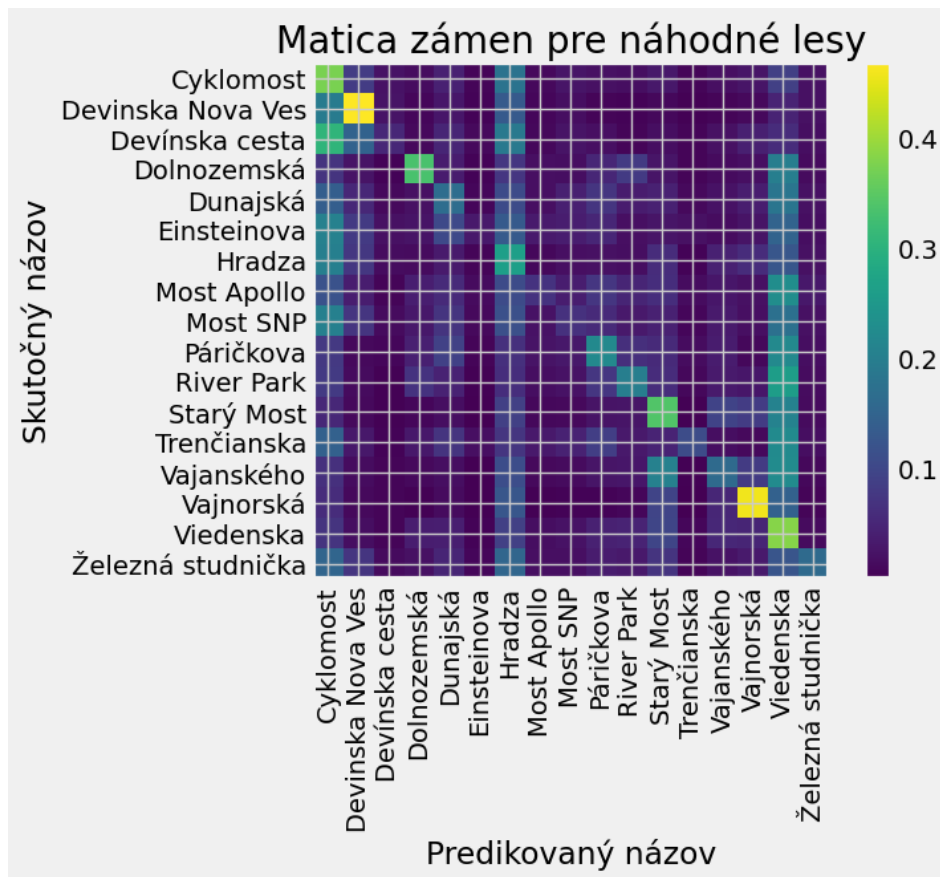


Obr. 4.1: Matica zámen pre logistickú regresiu bez hodinových okien.

4.2.2 Random Forest Classifier

Aby sme zachytili zložité vzťahy v dátach, použili sme aj algoritmus náhodného lesa. Použili sme rovnaké údaje ako pri logistickej regresii, aby sme mohli porovnať výsledky. Síce sa presnosť modelu

zvýšila na 29%, model stále nedokázal spoľahlivo rozlíšiť medzi trasami, čo ukázalo potrebu prejsť na analýzu časových radov. Na Obrázku 4.2 je možné pozorovať jemnú diagonalizáciu matice zámen, ale stále vidno značné ‘svetlejšie’ vertikálne čiary nad *Viedenskou* / *Hrádzou* (zberané už 11 rokov), zatiaľ čo nad *Einsteinovou* (zberané približne rok) je ‘tmavo’.



Obr. 4.2: Matica zámen pre náhodné lesy bez hodinových okien.

4.2.3 Neurónová sieť (LSTM)

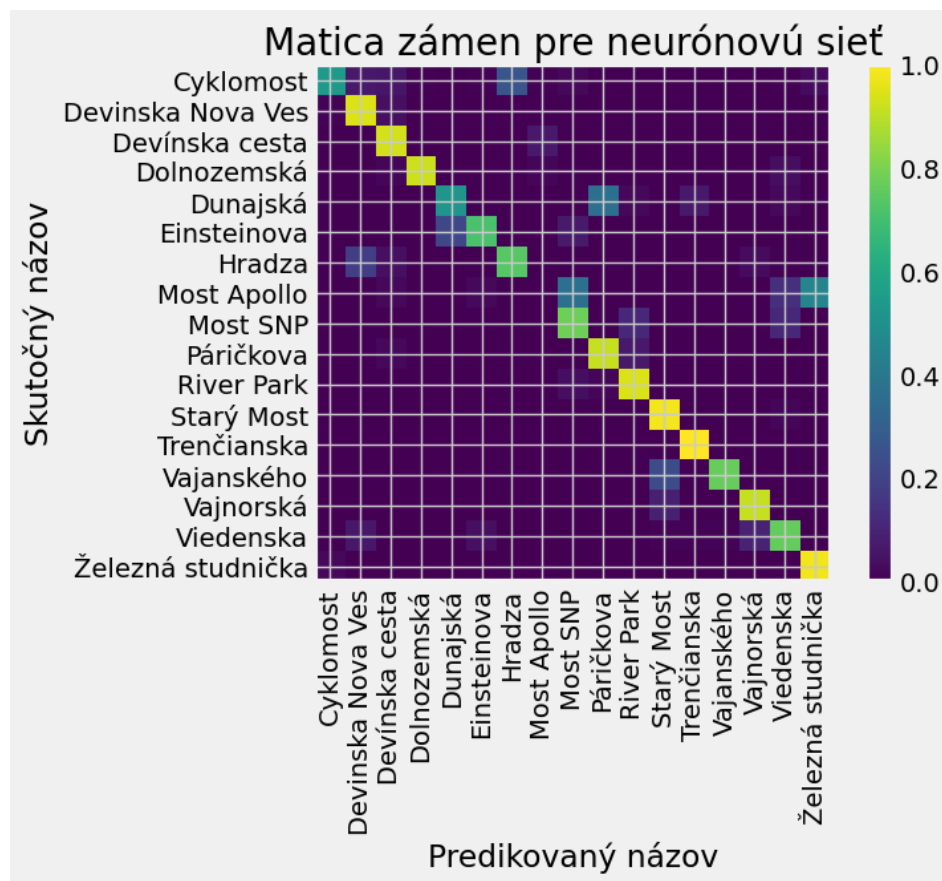
V treťom kroku sme prešli na hlboké učenie a využili sme predspracované týždenné okná. Model bol vytvorený pomocou Keras/TensorFlow a obsahoval tieto časti:

- **Bidirectional LSTM:** Hlavná časť modelu - učí sa dlhodobé vzťahy v časových radoch a spracováva sekvenciu v oboch smeroch.
- **Dropout:** Technika, ktorá počas tréningu náhodne vypína niektoré neuróny, aby sme predišli preučeniu a donútili model učiť sa lepšie vzory.

4.3 Výsledky klasifikácie a vyhodnotenie

Po transformácii dát na týždenné časové okná a natrénovaní modelov sme pristúpili k vyhodnoteniam ich úspešnosti. Zatiaľ čo celková presnosť nám poskytuje rýchly prehľad, analýza chybovosti pomocou

matice zámen nám pomáha s identifikáciou vzorcov chybovosti. Na nasledujúcom obrázku 4.3 môžeme pozorovať, ako model klasifikoval jednotlivé testovacie vzorky oproti ich skutočným označeniam.



Obr. 4.3: Matica zámen pre neurónovú sieť s hodinovými oknami.

Z tejto vizualizácie možno model hodnotiť pozitívne, pretože ‘svetlé’ štvorčeky sú prevažne na diagonále, čo znamená lepšiu presnosť.

Detailnejší pohľad na výkonnosť pre jednotlivé triedy metriky Precision (presnosť), Recall (návravnosť) a F1-skóre. Tieto hodnoty sú kľúčové najmä kvôli spomínanej nevyváženosti datasetu, kde dominantné triedy (trasy s dlhou históriou) môžu skresľovať celkovú úspešnosť.

Cyklotrasa	Precision	Recall	F1-score	Support
Cyklomost	97.96	53.63	69.31	179
Devínska Nová Ves	62.33	94.79	75.21	96
Devínska cesta	62.69	93.33	75.00	45
Dolnozemska	97.96	92.31	95.05	52
Dunajská	88.46	52.27	65.71	44
Einsteinova	52.63	71.43	60.61	14
Hradza	73.14	73.99	73.56	173
Most Apollo	0.00	0.00	0.00	44
Most SNP	60.34	77.78	67.96	45
Párikova	70.69	91.11	79.61	45
River Park	84.48	94.23	89.09	52
Starý Most	79.69	98.08	87.93	52
Trenčianska	73.68	100.00	84.85	14
Vajanského	92.00	76.67	83.64	30
Vajnorská	65.08	91.11	75.93	45
Viedenska	88.59	76.30	81.99	173
Železná studnička	63.24	97.73	76.79	44
accuracy			75.59	1147
macro avg	71.35	78.52	73.07	1147
weighted avg	76.70	75.59	74.05	1147

Tabuľka 4.1: Výsledky klasifikácie pre jednotlivé cyklotrasy neurónovej siete s hodinovými oknami.

Z tabuľky 4.1 môžeme vyčítať, že model dosahuje vysokú úspešnosť pri trasách, ktoré majú špecifický a nezameniteľný profil (napr. výrazné rekreačné špičky cez víkend), zatiaľ čo pri generických mestských trasách s bežnou dochádzkovou špičkou je presnosť nižšia (napr. *Most Apollo*, *Most SNP*, *Einsteinova*).

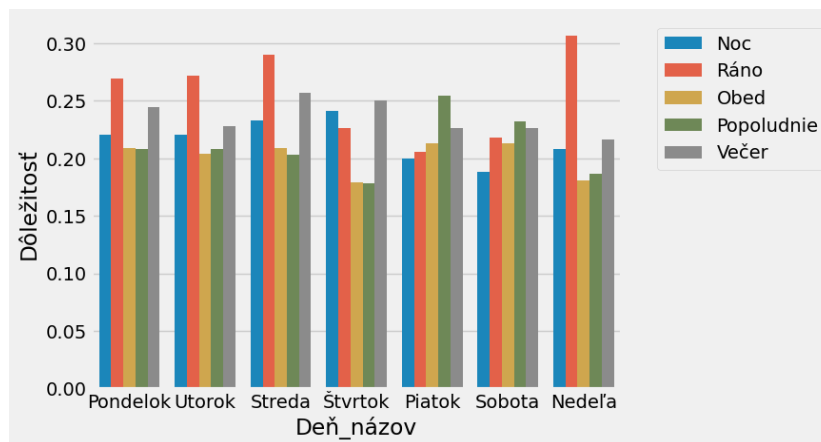
Pri neurónových sieťach nie je výstupom len jedna trieda, ale vektor pravdepodobností pre všetky možné cyklotrasy. To nám umožnilo analyzovať nielen to, či sa model ‘trafil’, ale aj to, ako blízko bol k pravde v prípade omylu. Vykonali sme analýzu chybné klasifikovaných vzoriek, aby sme zistili, či sa správna odpoveď nachádzala aspoň medzi 2. alebo 3. najpravdepodobnejšou voľbou modelu. Celkový počet nesprávnych klasifikácií bol 280, z tých tá správna odpoveď bola na 2. mieste v 121 klasifikáciách a na 3. mieste v 65 klasifikáciách. Tento výsledok naznačuje, že aj keď model neurčil presnú lokalitu ako víťaznú, v mnohých prípadoch ju považoval za veľmi pravdepodobnú alternatívu. To potvrdzuje hypotézu, že v Bratislave existujú skupiny cyklotrás s takmer identickými vzorcami správania cyklistov, ktoré je ťažké rozlíšiť len na základe historických dát bez ďalšieho kontextu.

4.4 Analýza dôležitosti atribútov

Hoci neurónové siete (LSTM) dosiahli v klasifikácii najlepšie výsledky, ich nevýhodou je nízka interpretovateľnosť (tzv. „black-box“ modely). Z ich vnútornej štruktúry je náročné určiť, ktorý konkrétny čas alebo deň rozhodol o zaradení trasy.

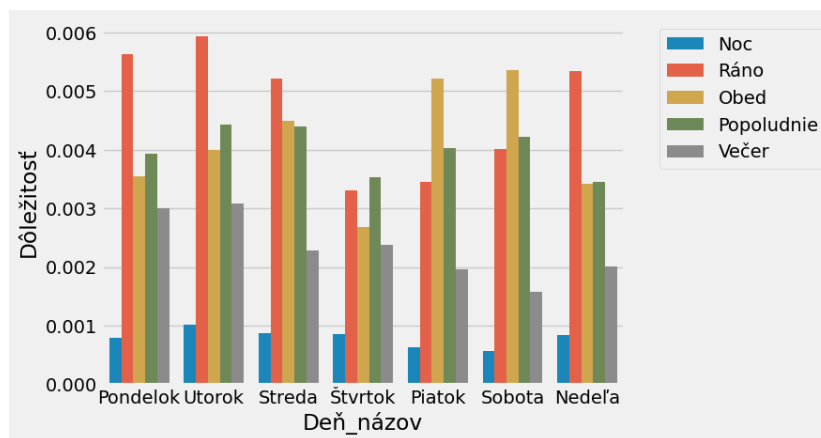
Aby sme lepšie pochopili, ktoré časové okná nesú najviac informácie pre rozlíšenie cyklotrás, vrátili sme sa k analýze dôležitosti atribútov pri modeloch *logistickej regresii* a *náhodných lesov*. Dáta sme pre prehľadnosť zoskupili podľa dní v týždni a častí dňa (Noc, Ráno, Obed, Popoludnie, Večer).

Pri lineárnom modeli sa ukázalo na Obrázku 4.4, že koeficienty dôležitosti sú rozdelené chaoticky. Model nedokázal nájsť jeden dominantný časový úsek, ktorý by univerzálne slúžil na identifikáciu trasy.



Obr. 4.4: Rozloženie dôležitosti atribútov pri Logistickej regresii – bez jasného vzoru.

Na rozdiel od regresie, model náhodných lesov odhalil logické a jasne interpretovateľné vzory správania, ktoré pomáhajú klasifikovať trasy: **Noc** má zanedbateľný vplyv na klasifikáciu (nízky koeficient). Keďže v noci je aktivita minimálna na všetkých trasách rovnako, tento časový úsek nenesie žiadnu rozlišovaciu informáciu. **Ráno (Po – Str, Ned)**: Ranná špička je kľúčovým identifikátorom najmä v prvej polovici pracovného týždňa a prekvapivo aj v nedeľu (športová aktivita). **Obéd (Pia, So)**: S blížiacim sa víkendom sa ťažisko dôležitosti presúva. V piatok a sobotu je pre rozlíšenie trasy najpodstatnejšia aktivita okolo obeda. **Popoludnie** si zachováva relatívne vysokú dôležitosť počas celého týždňa, čo značí, že poobedné návraty z práce alebo výletov sú silným signálom pre model. Počas **víkendových večerov** dôležitosť opäť klesá.



Obr. 4.5: Dôležitosť atribútov pri modeli Random Forest – viditeľné trendy pre rána a obedy.

Úspešné modely a následná analýza ukázali, že cyklotrasy v Bratislave majú svoje unikátne časové profily. Zatiaľ čo jednoduché modely dokážu zachytiť základné trendy (napríklad rozdiel medzi pracovným ránom a víkendovým obedom), pre presnú klasifikáciu konkrétnej lokality je nevyhnutné použiť komplexnejšie modely neurónových sietí, ktoré vnímajú týždeň ako spojitú sekvenciu.

Kapitola 5

Analýza vplyvu novej cyklotrasy na Vajanského nábreží

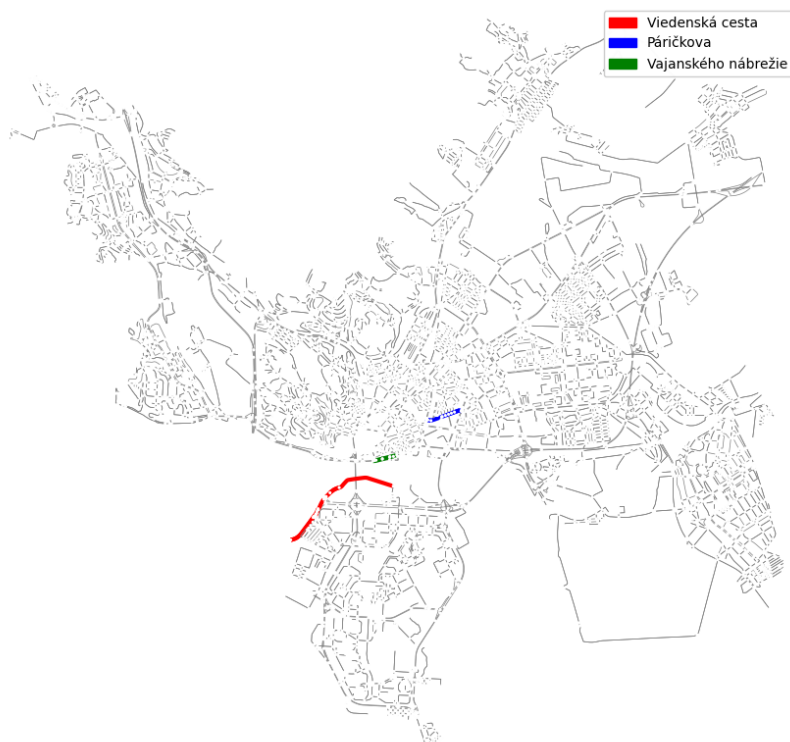
V tejto kapitole budeme analyzovať vplyv cyklotrasy na Vajanského nábreží v Bratislave, ktorá bola otvorená 1. septembra 2023. Táto trasa spája mestské centrum s rekreačnými oblasťami a zároveň poskytuje alternatívnu cestu pre dochádzajúcich cyklistov.

Cieľom tejto kapitoly je kvantifikovať kauzálny vplyv otvorenia cyklotrasy na Vajanského nábreží na intenzitu cyklistickej dopravy a identifikovať charakter jej využitia. Konkrétne sa zameriavame na odpovede na nasledujúce výskumné otázky: Viedlo otvorenie cyklotrasy k štatisticky signifikantnému nárastu počtu cyklistov? Aký je časový profil tohto vplyvu? Jedná sa o okamžitý skok alebo postupný nárast? Aké sú charakteristiky využitia novej trasy v porovnaní s typickými dochádzkovými a rekreačnými trasami v meste?

5.1 Dáta

Na cyklotrase na Vajanského nábreží sú nainštalované dva paralelné sčítače (Vajanského 1 a Vajanského 2). Keďže tieto sčítače monitorujú tú istú infraštruktúru, ich hodnoty boli v každom časovom okamihu sčítané a ďalej analyzované ako jedna trasa „Vajanského“. Sčítače na Vajanského nábreží poskytujú dáta až od konca januára 2024. Do osadenia nových sčítačov na Vajanského, mesto vychádzalo z čísel cyklosčítačov pri River Parku. Pre zachytenie vývoja cyklistickej dopravy pred týmto obdobím bola využitá lokalita River Park, ktorá sa nachádza v bezprostrednej blízkosti analyzovanej trasy a vykazuje dlhodobú dostupnosť dát už od roku 2022.

Na porovnanie správania cyklistov na rôznych typoch trás boli do analýzy zahrnuté aj ďalšie dve lokality: **Pářičkova**, reprezentujúca predovšetkým dochádzkovú cyklistickú dopravu a **Viedenská**, reprezentujúca rekreačnú cyklistickú trasu. Meteorologické dáta boli získané zo služby Meteostat pre lokalitu Bratislava. Použité premenné zahŕňajú priemernú dennú teplotu, úhrn zrážok a priemernú rýchlosť vetra.



Obr. 5.1: Prehľad analyzovaných cyklotrás

5.1.1 Čistenie a príprava dát

Pôvodné dáta zo sčítačov sú dostupné v hodinových intervaloch a rozlišujú smer jazdy. Pre každý záznam bol preto najskôr vypočítaný celkový počet cyklistov ako súčet prejazdov v oboch smeroch. Následne boli dáta agregované na dennú úroveň, čo umožňuje eliminovať krátkodobé fluktuácie v rámci dňa a zároveň zabezpečuje kompatibilitu s dennými meteorologickými údajmi.

V prípade Vajanského nábrežia boli dáta z oboch paralelných sčítačov transformované do spoločného časového formátu a následne sčítané, čím vznikol jednotný denný časový rad reprezentujúci intenzitu cyklistickej dopravy na tejto trase. Keďže sčítače na Vajanského nábreží sú dostupné až od konca januára 2024, pre obdobie pred ich inštaláciou bol, ako referenčná lokalita, použitý sčítač River Park. Denné dáta z River Parku boli použité výlučne do momentu, keď sa začínajú objavovať dáta zo sčítačov na Vajanského nábreží, aby sa predišlo prekryvaniu rôznych zdrojov merania.

Takto skonštruovaný časový rad umožňuje analyzovať vývoj cyklistickej dopravy na Vajanského nábreží v dlhšom časovom horizonte, pričom zmena meracej lokality nevedie k diskontinuite v dátach. Predpokladom tohto prístupu je, že River Park a Vajanského nábrežie zachytávajú podobné dopravné vzorce, keďže ide o priestorovo blízke úseky tej istej dopravnej osi.

Dátum otvorenia cyklotrasy na Vajanského nábreží bol stanovený na 1. september 2023. Na identifikáciu potenciálneho vplyvu tejto udalosti bola vytvorená binárna premenná, ktorá nadobúda hodnotu 1 v období po otvorení cyklotrasy a 0 v období pred ním. Okrem toho bola skonštruovaná trendová premenná vyjadrujúca počet dní od otvorenia cyklotrasy, ktorá umožňuje zachytiť postupnú adaptáciu cyklistov na novú infraštruktúru.

Meteorologické údaje boli časovo zosúladené s cyklistickými dátami na dennej báze a spojené pomocou dátumu. Do výsledného datasetu boli zahrnuté priemerná denná teplota, úhrn zrážok a priemerná

rýchlosť vetra, ktoré predstavujú kľúčové faktory ovplyvňujúce rozhodnutie o využití bicykla.

Na kontrolu rozdielov medzi pracovnými dňami a víkendmi bola vytvorená binárna premenná, ktorá nadobúda hodnotu 1 pre soboty a nedele a 0 pre pracovné dni. Táto premenná umožňuje oddeliť rekreačné cyklistické správanie od každodennej dochádzky.

Výsledkom procesu čistenia a prípravy dát je konzistentný denný časový rad, ktorý kombinuje dáta z viacerých sčítačov, explicitne zohľadňuje otvorenie cyklotrasy a zároveň kontroluje vplyv základných meteorologických a kalendárnych faktorov.

5.2 Regresný model

Analýza bola realizovaná pomocou lineárnych regresných modelov, odhadovaných metódou najmenších štvorcov na denných agregovaných dátach. Keďže analyzované dáta majú charakter denného časového radu, nie je splnený predpoklad nezávislosti a konštantného rozptylu rezíduí vyžadovaný klasickým OLS modelom. Pri dátach o cyklistickej doprave možno očakávať autokoreláciu medzi po sebe nasledujúcimi dňami aj rozptyl rezíduí spôsobený sezónnosťou. Z tohto dôvodu boli použité HAC smerodajné chyby podľa Newey–West, ktoré poskytujú konzistentné odhady aj v prítomnosti autokorelácie a rozptyl neznámeho tvaru. Počet oneskorení bol nastavený na 7 dní, aby zachytil týždenný cyklus v správaní cyklistov.

5.2.1 Model pre trasu Vajanského nábrežie

Pre trasu Vajanského nábrežie bol použitý model s prerušovaným trendom, ktorého cieľom je zachytiť zmenu vývoja cyklistickej dopravy po otvorení novej cyklistickej infraštruktúry v septembri 2023. Model zahŕňa:

- lineárny časový trend (`days_since_start`),
- binárnu premennú indikujúcu obdobie po otvorení cyklotrasy (`post_opening`),
- premennú umožňujúcu zmenu sklonu trendu po otvorení (`days_after_opening`),
- meteorologické kontrolné premenné (teplota, zrážky, rýchlosť vetra),
- indikátor víkendových dní,
- mesačné indikátory na kontrolu sezónnosti.

5.2.2 Porovnávacie modely pre trasy Páříčkova a Viedenská

Pre trasy Páříčkova (dochádzková trasa) a Viedenská (rekreačná trasa) bol použitý jednoduchší regresný model bez časového trendu a bez intervencie, keďže na týchto trasách nedošlo k porovnateľnej infraštruktúrnej zmene v analyzovanom období.

Vysvetľujúce premenné v týchto modeloch zahŕňajú:

- meteorologické kontrolné premenné (teplota, zrážky, rýchlosť vetra),
- indikátor víkendových dní,
- mesačné indikátory na kontrolu sezónnosti.

Tieto modely slúžia primárne na porovnanie citlivosti rôznych typov cyklistických trás na počasie a sezónnosť, nie na identifikáciu kauzálneho efektu infraštruktúrnej zmeny.

5.3 Výsledky

Kľúčovým zistením analýzy je štatisticky signifikantný pozitívny efekt premennej `days_after_opening`, zachytávajúcej zmenu trendu po otvorení cyklotrasy. Odhadnutý koeficient tejto premennej dosahuje hodnotu 0.912 s p-hodnotou $p = 0.039$, čo indikuje, že po otvorení novej infraštruktúry v septembri 2023 sa denný počet cyklistov zvyšuje v priemere o približne jedného cyklistu za každý deň. Tento postupný nárast naznačuje, že cyklisti si novo vybudovanú trasu osvojujú postupne v čase. Kumulatívny efekt tohto postupného nárastu je podstatný, po sto dňoch od otvorenia predstavuje tento trend približne 91 dodatočných cyklistov denne v porovnaní s obdobím bezprostredne po otvorení.

Binárna premenná `post_opening`, indikujúca obdobie po otvorení cyklotrasy, vykazuje pozitívny koeficient 181.9, avšak tento efekt je na hranici štatistickej signifikancie s p-hodnotou $p = 0.079$. Tento výsledok možno interpretovať ako okamžitý nárast využitia bezprostredne po otvorení cyklotrasy, hoci je potrebné pristupovať k tomuto zisteniu s opatrnosťou vzhľadom na marginálnu štatistickú signifikanciu.

Kombinácia oboch efektov, okamžitého nárastu a postupného trendu, naznačuje, že otvorenie cyklotrasy malo dvojázový vplyv na cyklistickú dopravu, kde sa počiatočný skok v návštevnosti kombinoval s následným postupným nárastom využitia.

Priemerná denná teplota má na Vajanského nábreží výrazný pozitívny efekt. Zvýšenie teploty o 1 °C je spojené s nárastom približne o 43 cyklistických prejazdov denne, pričom tento efekt je vysoko štatisticky významný. Naopak, zrážky a vietor pôsobia negatívne. Každý milimeter zrážok znižuje denný počet cyklistov približne o 35, zatiaľ čo zvýšenie rýchlosti vetra o 1 m/s vedie k poklesu o približne 13 prejazdov denne.

Mesačné indikátory zachytávajú výraznú sezónnosť v cyklistickej doprave. Najvyššia aktivita sa pozoruje v jarných a letných mesiacoch, pričom máj a jún vykazujú najvyššie koeficienty okolo 400 cyklistov v porovnaní s januárom ako referenčným mesiacom. Jesenné a zimné mesiace vykazujú podstatne nižšiu aktivitu, pričom december dokonca vykazuje negatívny koeficient v porovnaní s januárom, hoci tento efekt nie je štatisticky signifikantný.

Zaujímavým špecifikom Vajanského nábrežia je jeho hybridný charakter. Analýza neukázala štatisticky významný rozdiel medzi pracovnými dňami a víkendmi, čo znamená, že trasa je rovnomerne využívaná počas celého týždňa a slúži teda rovnako intenzívne na dochádzanie do práce, ako aj na voľnočasové a rekreačné jazdy.

5.3.1 Porovnanie s dochádzkovými a rekreačnými trasami

Pre kontextualizáciu výsledkov z Vajanského nábrežia boli analyzované aj dve ďalšie trasy s odlišným charakterom využitia: Páříčkova ako typická dochádzková trasa a Viedenská ako rekreačná trasa. Tieto porovnávacie analýzy umožňujú lepšie pochopiť špecifiká novej cyklotrasy na Vajanského nábreží.

Kľúčovým rozlišovacím znakom pri modeli pre trasu Páříčkova je výrazne negatívny koeficient víkendového indikátora na úrovni -203.66 , ktorý je vysoko štatisticky signifikantný. Tento pokles o viac než 200 cyklistov počas víkendov v porovnaní s pracovnými dňami predstavuje takmer 60-percentný pokles aktivity a jasne potvrdzuje primárne dochádzkovú funkciu tejto infraštruktúry. V kontraste s týmto výsledkom Vajanského nábrežia nevykazuje štatisticky signifikantný rozdiel medzi pracovnými dňami a víkendmi, čo naznačuje jeho zmiešanú funkciu.

Víkendový indikátor modelu pre Viedenskú trasu má pozitívny koeficient 198.19, ktorý je vysoko štatisticky signifikantný s prakticky nulovou p-hodnotou. Tento nárast takmer 200 cyklistov počas víkendov v porovnaní s pracovnými dňami predstavuje približne 84-percentný nárast oproti baseline úrovni a jednoznačne potvrdzuje rekreačný charakter tejto trasy. Tento výsledok stojí v dokonalom kontraste s Páříčkovou, kde víkendy znamenajú pokles aktivity, zatiaľ čo na Viedenskej predstavujú vrchol využitia.

Na Párickovej koeficient teploty dosahuje hodnotu 22.87, čo pri baseline úrovni približne 341 cyklistov predstavuje relatívny nárast približne 6.7 % pri zvýšení teploty o jeden stupeň Celzia. Na Viedenskej je koeficient teploty 33.05, čo pri baseline úrovni približne 236 cyklistov predstavuje relatívny nárast približne 14 %. Podobný vzorec sa pozoruje aj pri zrážkach, kde Viedenská vykazuje pokles približne 11 % na milimeter zrážok, v porovnaní s 5.6 % na Párickovej. Rýchlosť vetra má na Viedenskej taktiež výraznejší relatívny vplyv než na dochádzkovej trase.

Vajanského nábrežie vykazuje meteorologickú citlivosť, ktorá sa nachádza medzi týmito dvoma extrémami. Koeficient teploty 42.96 pri baseline úrovni približne 617 cyklistov predstavuje relatívny nárast približne 7 %, čo je porovnateľné s dochádzkovými trasami. Táto podobnosť v relatívnych efektoch naznačuje, že hoci Vajanského nábrežie má vyššiu absolútnu baseline úroveň využitia, citlivosť používateľov na poveternostné podmienky je bližšie dochádzkovému než rekreačnému správaniu. V kombinácii s absenciou štatisticky signifikantného víkendového efektu to naznačuje, že trasa slúži predovšetkým pravidelným používateľom, ktorí ju využívajú konzistentne počas celého týždňa.

Sezónne vzorce zachytené mesačnými indikátormi vykazujú na všetkých troch trasách podobný kvalitatívny priebeh s maximom v jarňách a letných mesiacoch, hoci s odlišnou intenzitou. Viedenská vykazuje najvýraznejšiu sezónnosť, zatiaľ čo Párickova a Vajanského nábrežie vykazujú mierne nižšiu sezónnu amplitúdu, čo opäť potvrdzuje, že pravidelní dochádzkoví cyklisti využívajú infraštruktúru počas celého roka, hoci s nižšou intenzitou v zimných mesiacoch.

Porovnanie týchto troch lokalít poskytuje dôležitý kontext pre interpretáciu výsledkov z Vajanského nábrežia. Absencia signifikantného víkendového efektu a podobná relatívna citlivosť na meteorologické podmienky ako na dochádzkovej trase Párickova naznačujú, že Vajanského nábrežie slúži ako hybridná trasa kombinujúca dochádzkovú a rekreačnú funkciu. Táto dualita využitia môže čiastočne vysvetľovať silný pozitívny trend po otvorení cyklotrasy. Nová infraštruktúra priláka nielen dochádzajúcich cyklistov hľadajúcich bezpečnejšiu a príjemnejšiu cestu do práce, ale aj rekreačných cyklistov využívajúcich atraktívne prostredie pozdĺž Dunaja. Schopnosť novej infraštruktúry osloviť obe skupiny používateľov môže byť kľúčovým faktorom jej úspechu a mohla by slúžiť ako model pre budúce projekty cyklistickej infraštruktúry v meste.