

Analýza dát zo cyklosčítačov v Bratislave

Tím: Jednotné programovacie družstvo a.s.

Marek Šugár

Fakulta matematiky, fyziky a informatiky
Univerzita Komenského v Bratislave
2026

Úvod

V posledných rokoch je z hľadiska urbánneho plánovania 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.

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 obyvateľov Bratislavy.

V tomto reporte je cieľom poukázať na zaujímavé interpretačné dôsledky, vychádzajúce z dát. Dôraz je kladený na interpretáciu dát s rôznou frekvenciou a spojitosti, resp. rozdielnosti medzi jednotlivými cyklotrasami. Okrem iného, hlbší dôraz je kladený na spojitosť celkovej vyťaženia cyklotrás s počasím.

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 Spojitosť vyťaženia cyklotrás a počasia	
Marek Šugár	3
1.1 Predspracovanie dát	3
1.2 Lineárna regresia	4
1.3 Interpretácia výsledkov	5
1.4 Penalizované lineárne regresie (LASSO, RIDGE)	7
2 Nejaká ďalšia kapitola	9

Kapitola 1

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.

1.1 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 spojitosti 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 ráme, takpovediac,

nutnosti.

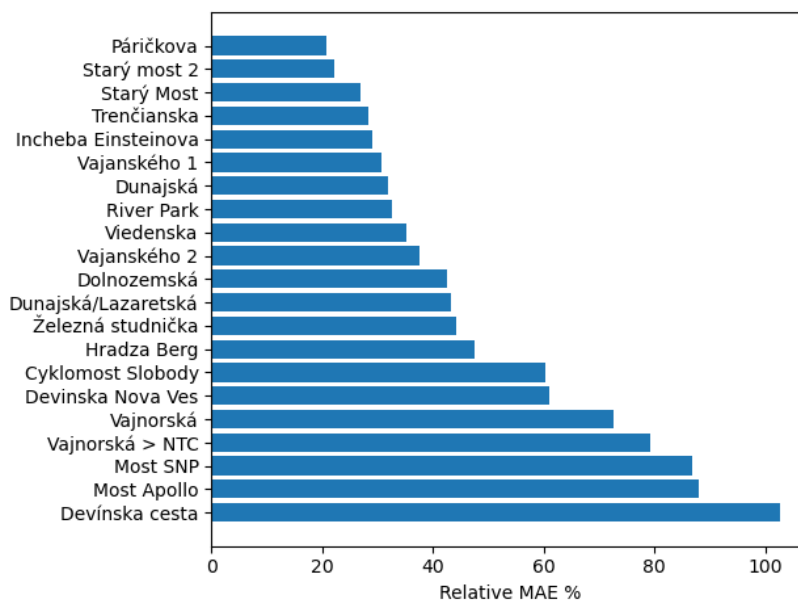
Pri delení dát do zodpovedajúcich tréningových, validačných a testovacích množín dát je nutné zachovať chronologickú následnosť dát, keďže v podstate sa jedná o časové rady, a neuváženým prístupom k deleniu, je možné disponovať skreslenými, zväčša nadhodnotenými, výsledkami výkonnosti modelov.

Z dôvodu variabilnej dĺžky monitorovaného časového intervalu pre rôzne cyklotrasy nie je možné pristupovať k delbe 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.

1.2 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 1.1.



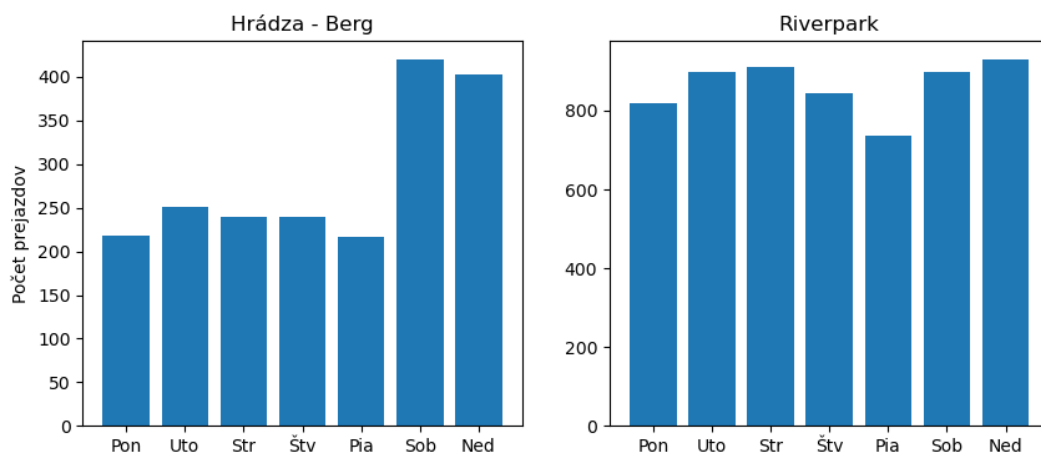
Obr. 1.1: Presnosť modelovania počtu prejazdov cyklotrasou lineárnou regresiou.

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%.

1.3 Interpretácia výsledkov

Na základe usporiadanej vizualizácie na Obrázku 1.1 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.

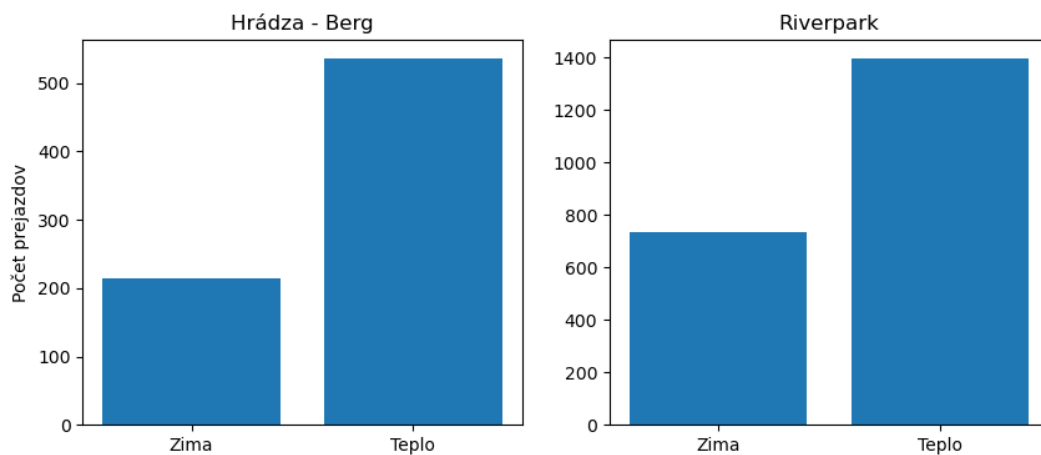
Táto skutočnosť je pomerne uchopiteľná. Cyklotrasy nachádzajúce sa v intenzívnejšie obývaných častiach mesta, pravdepodobne viac slúžia ako prostriedok dopravy do školy alebo zamestnania a teda môžu byť využívané takpovediac univerzálne – za každého počasia a približne v rovnakej miere počas týždňa. Opačne, periférnejšie cyklotrasy sú pravdepodobne využívané väčšinou rekreačne, pravdepodobne v rámci víkendov. Na Obrázku 1.2 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. 1.2: 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.

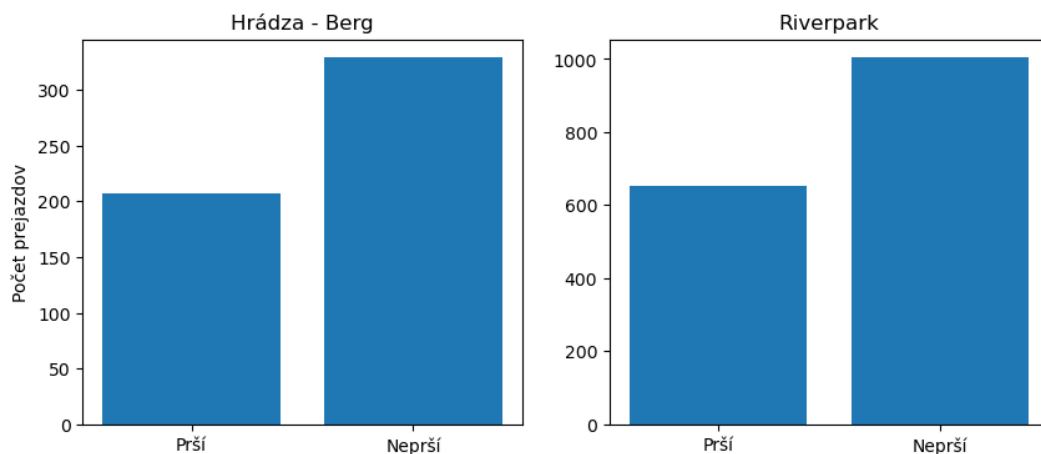
Táto skutočnosť je pomerne vhodne vizualizovaná na Obrázku 1.3, ktorý porovnáva priemernú využiteľnosť trás Riverpark a Hrádza - Berg v dňoch s teplotami $> 20^{\circ}\text{C}$ (Teplo) a $\leq 20^{\circ}\text{C}$ (Zima) počas dňa.



Obr. 1.3: 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 centrálnějších. 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 1.4 je vyobrazený priemerný počet prejazdov trasou v dňoch s nulovým a nenulovým spádom zrážok počas dňa.



Obr. 1.4: 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

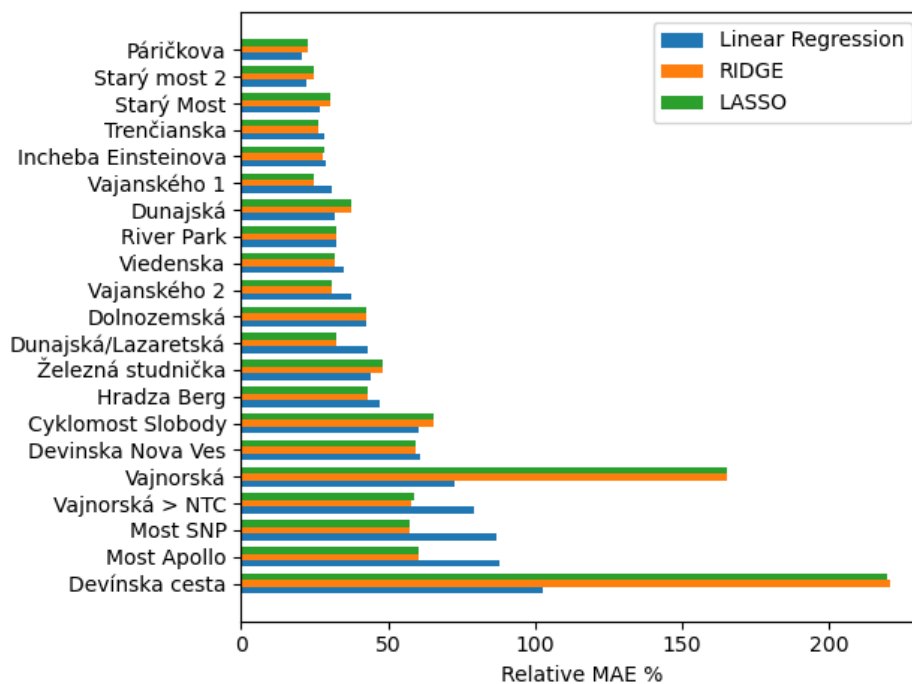
pozorujeme pokles na úrovni 36.96% a na trase Riverpark 34.89%. Napriek rovnakému usporiadaniu, čiže vyšší pokles pri Hradza – Berg, čo podporuje našu hypotézu, je rozdiel medzi jednotlivými trasami významne menší. To nás privádza k potenciálnemu použitiu modelovania s využitím istej formy výberu podvzorky premenných.

1.4 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 1.5 vyobrazujeme výslednú analýzu odchýlok pre jednotlivé regularizácie v porovnaní s lineárnou regresiou.

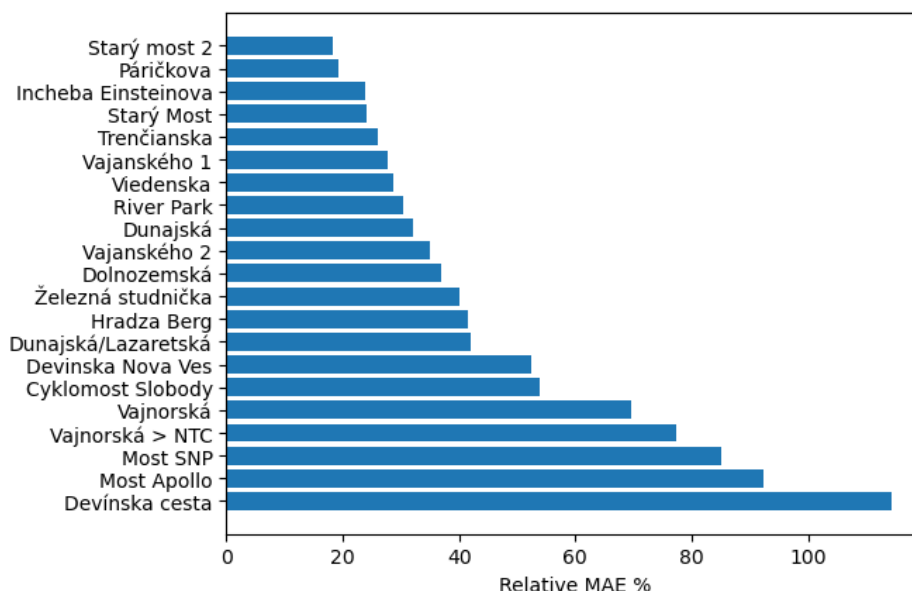


Obr. 1.5: Porovnanie výkonnosti predikcií počtu prejazdov na jednotlivých cyklocestách.

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 **Vaj-**

norská, kde oba typy regularizácie dosiahli obzvlášť vysokú odchýlku. Za spomenutie stojí aj to, že v drivej väčšine trás je miera odchýlok vo všetkých troch algoritmoch veľmi podobná.

Prírodnou 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 1.6.



Obr. 1.6: Výkonnosť predikcii pomocou **RandomForestRegression** na jednotlivých cyklotrasách.

Je možné pozorovať relatívne obdobné usporiadanie úspešností regresie ako na Obrázku 1.5. Akési rozdelenie provinčných a centrálnych cyklotrás v jednotlivých poloviciach je taktiež pomerne zjavné.

Záverom je možné konštatovať viaceré záverov. Prístupom trénovania 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 1.2 preukázali koncentrickosť cez víkendy. Tým sa ukázalo, že takéto porovnávanie 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**, **Hradza – 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 2

Nejaká ďalšia kapitola