

# KLASIFIKACE A ČASOVÁ ANALÝZA OSOBNÍ A NÁKLADNÍ VLAKOVÉ DOPRAVY REGIONÁLNÍCH TRATÍ

## Cíl práce

Hlavním cílem diplomové práce je provést klasifikaci a časovou analýzu provozu osobní a nákladní dopravy na vybraných celostátních a regionálních železničních tratích (bez železničních koridorů) Olomouckého a Moravskoslezského kraje. K dosažení cílů bude využita víceletá datová sada agregovaných měsíčních údajů o vlakové dopravě na tratích České republiky. Nejprve bude provedeno předzpracování dat, zahrnující selekci údajů pouze pro definované území, vytvoření prostorových vrstev pro mapové vizualizace a očištění dat od kalendářních vlivů. Poté budou pro vybrané úseky vytvořeny časové řady, na jejichž základě bude vyhodnocován provoz osobní a nákladní dopravy od roku 2016 do roku 2021. Nakonec bude provedena klasifikace metodou shlukování, spočívající v identifikaci tratí s podobným a současně odlišným provozem, a určení tratí s podobným časovým průběhem provozu.

Dalšími cíli práce je hodnocení vlivu dopadů pandemie covidu-19 na osobní a nákladní dopravu v letech 2020 a 2021. Dále zjišťování trendů v provozu na tratích pomocí dekompozice časových řad. V případě, že to bude realizovatelné, tak bude sledována i složka sezónní. Při shlukování budou využity a porovnány přístupy hierarchické metody a metody k-means. U hierarchické metody budou testovány různé metody spojování shluků a vzdálenostních metrik, z nichž bude nakonec vybrána nevhodnější kombinace pro konkrétní analyzovaná data. Posledním cílem je využití shlukování časových řad metodou dynamického borcení časové osy pro hledání tratí s podobným časovým průběhem.

## Výsledky

### STRUKTURA ŽELEZNIČNÍ DOPRAVY

Podle celkového počtu a hmotnosti vlakových souprav za rok 2016 a 2021 byl největší provoz zjištěn na tratích v okolí Ostravy. Největší provoz v letech 2016 a 2021 byl na trati Ostrava-Stodolná – Ostrava-Střed. Další tratě s vysokým provozem byly Šumperk–Zábřeh, Přerov–Nezamyslice, Olomouc–Prostějov, Opava–Ostrava, Hranice–Valašské Meziříčí. Naopak nejmenší provoz byl na regionálních tratích. Na drtivě většině všech tratí převažovala osobní doprava. Odhaleny byly tratě, na kterých probíhala pouze osobní nebo nákladní doprava.

Největší poklesy osobních vlaků o více než 50 % zaznamenaly v roce 2020 úseky na hraničních přechodech. Důvodem bylo na určitou dobu uzavření státních hranic pro osobní dopravu. Počty vlaků klesly především na regionálních tratích. Poklesy však zaznamenaly i provoz na celostátních tratích. Na některých úsecích došlo naopak k vyššímu počtu průjezdů vlaků. Celkově byl počet vypravených vlaků v březnu–květnu 2020 a 2021 nižší než v předchozích letech.

### ČASOVÁ ANALÝZA

Vyhodnocením časových řad měsíčních agregací byl popsán provoz na vybraných úsecích od roku 2016 do roku 2021. Pomocí časových řad byly odhaleny výkyvy v provozu a také především vliv pandemie na osobní a nákladní dopravu v letech 2020 a 2021. Na úsecích celostátních tratí byl obecně velký počet osobních vlakových souprav a provoz byl zatížen menšími výkyvy. Na úsecích celostátních tratí probíhá pravidelná nákladní doprava s velkým významem. Pandemie nejvíce postihla provoz na regionálních tratích, kdy v posledních dvou letech na těchto tratích docházelo k častým a výrazným propadům v osobní dopravě.

Na úsecích v blízkosti průmyslových areálů byla nákladní doprava ovlivněna především na jaře 2020. Nejvíce byl zasažen úsek v blízkosti automobilky Hyundai Nošovice. Tyto zjištění potvrdily utlumení průmyslové výroby na počátku šíření covidu-19 v ČR. Minimální dopady pandemie byly na tratě se specifickým provozem. Naopak výrazně byla ovlivněna doprava na hraničních přechodech. Na základě rozhodnutí o uzavření státních hranic neprojel hraničními přechody v dubnu a květnu 2020 žádný osobní vlak.

### KLASIFIKACE TRATÍ

Vyzkoušeny byly dva přístupy zpracování využitím hierarchického shlukování a shlukování metodou k-means. Pro obě metody byly sestaveny modely v programu Orange. Experimentováno bylo s různými kombinacemi metod, metrik a cílových počtů shluků. U hierarchického shlukování se nejvíce osvědčila Wardova metoda s využitím Mahalanobisovy nebo Euklidovy vzdálenosti. Specificky odlišný provoz byl odhalen na úsecích Bohumín–Vrbice st. hr. – Bohumín–Vrbice S618a a Bohumín–Vrbice – Bohumín–Vrbice S618. Úseky podobné nízkým provozem byly především na regionálních tratích. Rozdíly ve výsledcích hierarchického a k-means shlukování byly zanedbatelné. Klasifikovány byly tratě s podobným průběhem provozu v letech 2016–2019, tedy v době před pandemií covidu, v roce 2020 a poté 2021.

## Data a metody

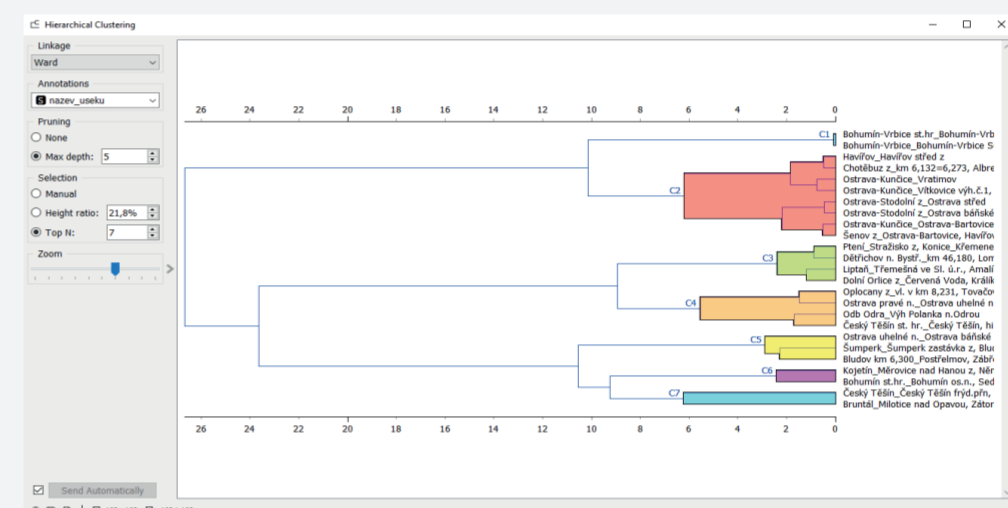
### POUŽITÁ DATA

Využita byla data provozu na železnici. Vlastníkem dat je státní organizace Správa železnic. Data obsahují informace o využití dráhy v měsíčních agregacích za jednotlivý rok podle kategorie dopravy. V datech se rozlišuje lokomotivní, nákladní, osobní, služební a soupravová kategorie železniční dopravy. Údaje o provozu jsou vztažena k monitorovacím bodům SR70. Monitorovací body se rozumí železniční stanice, dopravně zajímavá místa jako jsou například hradla či dopravní a tarifní místa. Jelikož se nejedná o data směrová, jde o úhrny provozu místa v obou směrech. Data obsahují popisné informace monitorovacího bodu, a především údaje o počtu vlaků, průměrné délce soupravy, hmotnosti souprav, průměrné hmotnosti soupravy, průměrném počtu náprav a průměrném počtu vozů, které daným bodem projely.

### METODY ZPRACOVÁNÍ

Časová analýza provozu byla založena na tvorbě časových řad měsíčních údajů počtu a hmotnosti vlaků za období 2016–2021. Časové řady byly vytvořeny v programu MS Excel. V programu Orange byla sestavena workflow, určená pro analýzu výsledných časových řad. Realizována byla dekompozice časových řad na trendovou a sezónní složku. Dále bylo sledováno meziroční tempo růstu počtu vlaků.

Klasifikace tratí byla realizována metodou shlukování. Při shlukové analýze železničních úseků byla využita především metoda hierarchického shlukování. Otestována byla taky k-středová metoda (k-means). Shlukování bylo realizováno v programu Orange. Pro interpretaci výsledných shluků, byly vytvořeny kategorizační tabulky podle počtu a hmotnosti vlakových souprav. Tyto tabulky umožnily klasifikovat shluky do 6 nadefinovaných kategorií. Každý shluk byl do příslušné kategorie přiřazen na základě hodnoty mediánu sledovaného atributu. Celková charakteristika shluků byla doplněna poznatkami získanými z dendrogramu v programu Orange, a také z mapových výstupů, které poskytly prostorové informace.



Obr. 1 Dendrogram hierarchického shlukování

