Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

POROVNÁNÍ SOUSEDNOSTI VYUŽITÍ ÚZEMÍ EVROPSKÝCH MĚST POMOCÍ FREKVENTOVANÝCH SAD

Diplomová práce

Bc. Pavel NOVÁK

Vedoucí práce Ing. Zdena Dobešová, Ph. D.

Olomouc 2023
Geoinformatika a kartografie

ANOTACE

Cílem diplomové práce bylo porovnat evropská města na základě frekventovaných sad vyjadřujících sousednost různých landuse (využití území) v evropských městech. Následně byl zahrnut popis charakterů měst, jejich porovnání a nalezení podobných měst. Dalším cílem byla realizace programu pro přípravu kategoriálních nebo dichotomických dat sousednosti polygonů využití území nutných pro generování frekventovaných sad. Práce využila zdrojová data využití území z Copernicus Urban Atlas 2018.

Diplomová práce implementuje netradiční metodu, běžně nepoužívanou v rámci geoinformatiky, a primárním výsledkem je nalezení postupů pro aplikování této metody. Teoretická část práce se zaobírala obecnými koncepty Data Miningu, popisem algoritmů pro výpočet frekventovaných sad a popisem studií, které více či méně implementují metodu frekventovaných sad na prostorová data.

V rámci práce byl vytvořen nástroj pro generování transakčních dat sousednosti využití území a provedeny tři případové studie na českých městech, evropských městech a porovnání výsledků se studií (Dobesova 2020). Pro česká města byly identifikovány signifikantní frekventované sady a pro evropská města byly pomocí hierarchického shlukování nalezeny skupin měst. Výsledky práce představují nový přístup v geoinformatice a mohou být využity pro porovnávání měst a identifikaci podobností a rozdílů mezi národními státy, popřípadě mezi městy.

KLÍČOVÁ SLOVA

Využití území, frekventované sady, sousednost, podobnost, evropská města

Počet stran práce: 91

ANOTATION

The master thesis aimed to compare European cities based on Frequent Itemsets expressing the adjacency of different land uses in European cities. Subsequently, a description of the character of the cities was included, comparing them, and finding similar cities. Another objective was to implement a program for the preparation of categorical or dichotomous land use polygon adjacency data necessary for the generation of Frequent Itemsets. The work used the source land use data from the Copernicus Urban Atlas 2018.

The thesis implements a non-traditional method, not commonly used within geoinformatics, and the primary outcome is to find procedures for applying this method. The theoretical part of the thesis dealt with general concepts of Data Mining, a description of algorithms for computing frequent itemsets and a description of studies that implement the Frequent Itemset method on spatial data.

The thesis developed a tool for generating transactional land use adjacency data and conducted three case studies on Czech cities and European cities and compared the results with the study (Dobesova 2020). Significant Frequent Itemsets were identified for Czech cities and city groups were found for European cities using hierarchical clustering. The results of the work represent a new approach in geoinformatics and can be used for city comparisons and identification of similarities and differences between nation-states or between cities.

KEYWORDS

Land use, Frequent Itemsets, Adjecency, Similarity, European Cities

Number of pages 91

Prohlašuji, že

- diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.

- jsem si vědom, že na moji diplomovou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,

- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užívat (§ 35 odst. 3),

- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

Poděkování patří především vedoucí práce doc. Ing. Zdeně Dobešové, Ph.D., která formou cenných rad a podnětů významně dopomohla k jejímu dokončení.

Dále bych chtěl poděkovat své přítelkyni Pavlíně Kostelníkové, která byla mou největší podporou a bez níž by tato práce nevznikla.

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2019/2020

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	**Bc. Pavel NOVÁK**
Osobní číslo:	R220649
Studijní program:	**N0532A330009 Geoinformatika a kartografie**
Téma práce:	**Porovnání sousednosti využití území evropských měst pomocí frekventovaných sad**
Zadávající katedra:	**Katedra geoinformatiky**

Zásady pro vypracování

Cílem práce je porovnat evropská města na základě frekventovaných sad vyjadřující sousednost různých landuse (využití území) ve městech. Student realizuje program na přípravu kategoriálních nebo dichotomických dat sousednosti polygonů nutných pro generování frekventovaných sad. Na základě zjištěných frekventovaných sad popíše charakter měst, jejich porovnání a nalezení podobných měst, či výjimek. Zdrojová data využití území budou použita z Copernicus Urban Atlas.

Celá práce (text, přílohy, výstupy, zdrojová a vytvořená data) se odevzdá v digitální podobě na paměťovém nosiči (CD, DVD, SD karta, flash disk). Text práce s vybranými přílohami bude odevzdán ve dvou svázaných výtiscích na sekretariát katedry. O diplomové práci student vytvoří webovou stránku v souladu s pravidly dostupnými na stránkách katedry. Práce bude zpracována podle zásad dle Voženílek (2002) a závazné šablony pro diplomové práce na KGI. Povinnou přílohou práce bude poster formátu A2.

Rozsah pracovní zprávy:	**max. 50 stran**
Rozsah grafických prací:	**dle potřeby**
Forma zpracování diplomové práce:	**tištěná**

Seznam doporučené literatury:

Bučková S. Aplikace vyhledávání kolokačních vzorů na prostorová data, UP, diplomová práce, 2022

Biolab: Orange Visual Programming Documentation, Release 3, 2021, https://orange.biolab.si

Ester M., Kriegel H.P., Sander J. Algorithms and Applications for Spatial Data Mining, Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS, Taylor and Francis, 2001

Janoušek, M. Porovnání urbánního prostoru pomocí kruhových výsečí, Univerzita Palackého v Olomouci, Katedra geoinformatiky PřF, diplomová práce, 2019

Salmenkivi, M. Frequent Itemset Discovery. In: Shekhar S., Xiong H., Zhou X. (eds) Encyclopedia of GIS. Springer, Cham. 2017, https://doi.org/10.1007/978-3-319-17885-1_432

Urbančík F. Podobnost evropských měst a jejich funkčních území, UP, diplomová práce, 2022

Voženílek, V. Diplomové práce z geoinformatiky. Olomouc, Univerzita Palackého v Olomouci, 2002

Vedoucí diplomové práce:	**doc. Ing. Zdena Dobešová, Ph.D.** Katedra geoinformatiky

Datum zadání diplomové práce: 7. května 2020
Termín odevzdání diplomové práce: **5. května 2023**

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA GEOINFORMATIKY
17. listopadu 50, 771 46 Olomouc
-1-

L.S.

doc. RNDr. Martin Kubala, Ph.D.
děkan

prof. RNDr. Vít Voženílek, CSc.
vedoucí katedry

V Olomouci dne 16. prosince 2021

OBSAH

SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Význam
FI	Frequent Itemset
FIM	Frequent Itemset Mining (dolování frekventovaných sad)
FS	frekventovaná sada
FUA	Functional Urban Area (funkční městská/urbánní oblast)
GIS	geografický informační systém
GUI	Graphical User Interface
KDD	Knowledge Discovery in Databases (objevení znalostí v databázích)
MBA	Market Basket Analysis (analýza nákupního košíku)
SPMF	Sequential Pattern Mining Framework
UA	Urban Atlas

ÚVOD

Diplomová práce se zaměřuje na porovnání evropských měst na základě frekventovaných sad, které vyjadřují sousednost různých využití území v městech. Cílem práce je identifikovat charakteristické vlastnosti měst a nalézt podobnosti mezi nimi. K tomu je použita metoda frekventovaných sad, která je netradiční metodou v oblasti geoinformatiky.

Frekventované sady jsou klíčovým nástrojem v oblasti dolování dat a hledání asociačních pravidel. Tyto sady jsou definovány jako soubory položek, které se vyskytují společně v datasetu s frekvencí překračující předem stanovený práh, tzv. minimální podporu. Podpora souboru položek je důležitým faktorem při určování frekvence výskytu těchto souborů v celkovém počtu transakcí. V praxi se frekventované sady využívají například v oblasti marketingu k identifikaci nákupních vzorců.

Praktická část práce se zaměřuje na implementaci programu na přípravu kategoriálních nebo dichotomických dat sousednosti polygonů, které jsou nutné pro generování frekventovaných sad. Zdrojová data využití území jsou použita z Copernicus Urban Atlas 2018. V rámci práce jsou také prezentovány tři případové studie, které ověřují aplikovatelnost metody na prostorová data. První případová studie se zabývá českými městy a poskytuje souhrnný popis využití území a identifikuje signifikantní frekventované sady. Druhá případová studie prozkoumává 100 evropských měst a na základě frekventovaných sad nalézá podobnosti mezi nimi. Třetí navazuje na již existující studie podobnosti evropských měst a snaží se potvrdit jejich výsledky.

Celkově lze říci, že práce přináší nové poznatky o aplikaci metody frekventovaných sad na prostorová data a popisuje postupy pro aplikaci této metody.

1 CÍLE PRÁCE

Cílem práce je porovnat evropská města na základě frekventovaných sad vyjadřující sousednost různých landuse (využití území) ve městech. Student realizuje program na přípravu kategoriálních nebo dichotomických dat sousednosti polygonů nutných pro generování frekventovaných sad. Na základě zjištěných frekventovaných sad popíše charakter měst, jejich porovnání a nalezení podobných měst, či výjimek. Zdrojová data využití území budou použita z Copernicus Urban Atlas 2018.

2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Data mining, také nazývaný dolování dat, je disciplína v oblasti informačních technologií, která se zabývá využíváním pokročilých algoritmů k identifikaci skrytých vzorců, vztahů a trendů v obrovských datových souborech. Český pojem dolování dat je potřeba chápat ve významu dolování znalostí z dat, které přináší nové poznatky, nikoliv hromadění (dolování) dat bez dalšího zpracování. Data mining umožňuje objevit informace, které by jinak zůstaly skryty a poskytuje možnost efektivnějšího využití těchto informací. Na počátku byla pro tuto oblast používána nejrůznější označení jako například: information harvesting, data archeology, data destilery (Berka 2003). Rozdíl oproti klasické statistice a metodám strojového učení je podle Berky v kladení důrazu na přípravu dat a na interpretaci výsledků.

2.1 Data mining

Určit první významný akademický článek o dolování dat je obtížné, jelikož koncept dolování dat byl studován a vyvíjen po několik desetiletí. Nicméně jeden z nejranějších a vlivných článků o dolování dat je "Knowledge Discovery in Databases: An Overview" (Frawley et al. 1992). V tomto článku autorů Frawley a Piatetsky-Shapiro byl představen koncept objevování znalostí v databázích (KDD – Knowledge Discovery in Databases), který definovali jako "**netriviální proces identifikace platných, nových, potenciálně užitečných a v konečném důsledku srozumitelných vzorců v datech.**" Dále diskutovali kroky, které jsou součástí procesu KDD, jako například čištění dat, integraci dat, výběr dat, transformaci dat, objevování vzorců, hodnocení vzorců a prezentaci znalostí. Článek rovněž zdůraznil výzvy a příležitosti spojené s KDD, včetně potřeby škálovatelných algoritmů, důležitosti odborné znalosti a potenciálních výhod použití technik KDD v různých oblastech, jako jsou marketing, finance a zdravotnictví.

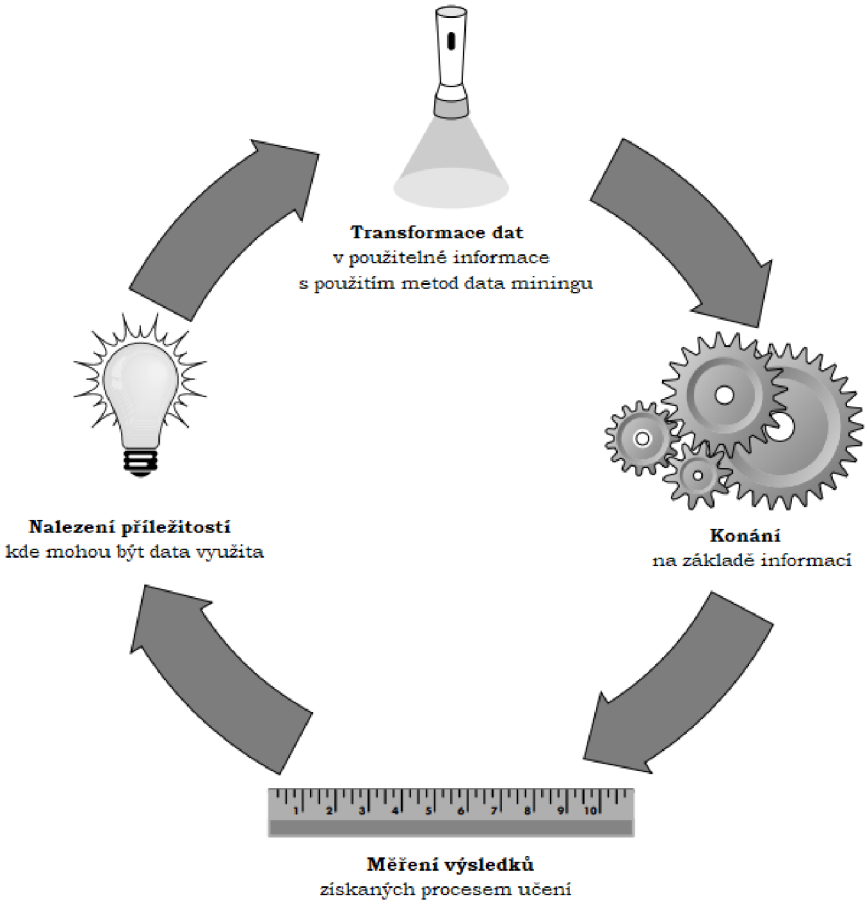
Podle jedné z prvních definic (Fayyad et al. 1996) lze data mining definovat jako "**proces objevování užitečných znalostí z velkých datových souborů**". Zahrnuje to využití pokročilých algoritmů k identifikaci trendů, vzorců a skrytých vztahů v datech. Autoři se zabývali procesem transformace surových dat na užitečné informace pomocí dolování dat a objevování znalostí v databázích. Zdůrazňují výzvy spojené s velkými objemy dat a nutnost použití sofistikovaných algoritmů a technik pro extrakci smysluplných poznatků. Popisují klíčové kroky v procesu KDD, včetně čištění dat, integrace dat, výběru dat, transformace dat, dolování dat, hodnocení vzorů a reprezentace znalostí. Podtrhují důležitost iterativního zlepšování v procesu KDD a také etická zvažování, která je třeba zohlednit. Celkově zdůrazňují transformační potenciál KDD v oblastech jako jsou obchod, zdravotnictví a vědecký výzkum (Fayyad et al. 1996).

Data mining se často používá k analýze velkých objemů dat z různých zdrojů. Data mining lze definovat také jako "**proces analýzy velkého množství dat z různých zdrojů s cílem najít v nich vzorce a vztahy, které mohou být využity pro předpovědi budoucích chování, tendencí a vztahů**" (Berry a Linoff 2004). Data mining také umožňuje identifikovat rizika a předpovědět výsledky pro podporu rozhodovacích procesů. Definice pochází z publikace: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, která přináší ucelený přehled o použití technik dolování dat v oblasti marketingu, prodeje a řízení vztahů se zákazníky. Obsahem je podrobný popis algoritmů a technik, které se používají pro analýzu dat a zjišťování skrytých informací, včetně shlukování, klasifikace, regrese a asociace. Dále ukazuje, jak se tyto techniky aplikují v praxi, například pro identifikaci zákaznických segmentů, predikci tržeb a optimalizaci marketingových kampaní. Autoři dále diskutují výzvy a rizika spojená

s použitím technik dolování dat v oblasti marketingu a vysvětluje základní zásady zabezpečení a ochrany soukromí dat.

Data mining umožňuje organizacím zlepšit svou výkonnost a konkurenceschopnost tím, že jim umožňuje analyzovat a využít data k dosažení efektivnějších výsledků. Data mining také pomáhá předvídat budoucí chování zákazníků a odhalit nové trhy a příležitosti. Autoři (Berry a Linoff 2004) definují **cyklus data miningu** na příkladu aplikace v ekonomické oblasti, ale tento cyklus lze zobecnit na celou problematiku dolování dat. Definují 4 fáze:

1. **Nalezení příležitosti**, kde mohou být data využita.
2. **Transformace dat** v použitelné informace s použitím metod data miningu.
3. **Konání** na základě informací.
4. **Měření výsledků** získaných procesem učení.



Obrázek 1 Cyklus data miningu – upraveno z (Berry a Linoff 2004).

V současné době existuje mnoho metod a technik pro aplikaci data mining, včetně klasifikačních algoritmů, shlukování, asociačních pravidel a regrese. Tyto metody se používají k identifikaci vzorců a vztahů v datech, které mohou být využity pro předpovědi a rozhodování. Význam data miningu je stále větší v důsledku velkého množství dat, která jsou k dispozici. V souvislosti s velkým množstvím dat nesmíme opomenout koncept **Big Data** (velká data), který se obecně odkazuje na obrovské množství dat, která jsou příliš velká, složitá a rychle se měnící, aby bylo možné je zpracovat tradičními metodami. V posledních letech se stala Big Data klíčovou součástí informačních technologií, které umožňují zpracovávat a analyzovat obrovské množství dat za účelem odhalení skrytých informací a vztahů.

Stěžejní metodou, která je v práci aplikována, je **metoda frekventovaných sad**. Prvním důležitým příspěvkem o frekventovaných sadách je "Fast Algorithms for Mining Association Rules", publikovaný autory R. Agrawal a R. Srikant v konferenčním sborníku 20. mezinárodní konference o velkých databázích (VLDB – Very Large Data Bases) v roce 1994. Autoři v tomto článku představili **Apriori algoritmus**, zásadní metodu pro objevování frekventovaných sad prvků v rozsáhlých transakčních databázích. Dále byl představen pojem **asociačních pravidel,** které jsou generovány na základě frekventovaných sad. "Fast Algorithms for Mining Association Rules" měl významný dopad na oblast dolování dat a inspiroval mnoho následných výzkumů a vývojů v oblasti dolování frekventovaných sad prvků a učení se asociačních pravidel.

2.2 Algoritmy pro získávání frekventovaných sad

Algoritmus Apriori je jeden z nejznámějších a nejpoužívanějších algoritmů pro dolování asociačních pravidel a frekventovaných sad v oblasti data miningu. Jeho základním principem je nalezení všech pravidel, která mají vysokou podporu v dané databázi transakcí. Hlavní myšlenkou algoritmu Apriori je, že každá podmnožina pravidla musí mít minimální podporu, aby pravidlo samo bylo zahrnuto mezi výsledná pravidla (Agrawal a Srikant 1994). Tento proces se opakuje postupně pro podmnožiny pravidla až do doby, kdy už neexistuje žádná podmnožina s minimální podporou a výsledná pravidla jsou tedy ty, která splňují stanovenou podporu a spolehlivost (Agrawal a Srikant 1994).

V oblasti data miningu má algoritmus Apriori velký význam, protože umožňuje identifikovat asociační pravidla a vztahy mezi různými položkami v datasetu. Používá se například pro doporučování produktů v obchodních aplikacích, analýzu zákaznických nákupních vzorců, segmentaci trhu a dalších oblastech, kde je potřeba analyzovat a identifikovat vzorce a vztahy v datech.

Jedním z nejznámějších a nejčastěji citovaných článků o algoritmu Apriori je práce Agrawal a Srikantové "Fast Algorithms for Mining Association Rules" z roku 1994, která popisuje originální verzi algoritmu Apriori a přináší detailní analýzu jeho složitosti a výkonu. Jak uvádí (Salmenkivi 2017) na algoritmus Apriori navázali další autoři, kteří přinesli komplexnější a efektivnější řešení (např. (Zaki 2000)).

Další z modifikací algoritmu Apriori je **algoritmus FP-growth**. Ten lze považovat za inovativní v tom, že pracuje s **kategoriálními daty** namísto **dichotomických dat**, jak je tomu u klasického Apriori algoritmu. Jako zastřešující pojem je používáno označení **transakční data** (Petr 2014a).

V rámci práce bylo navrženo využití **algoritmu negFIN** namísto klasického Apriori algoritmu. V rámci odborných konzultací bylo navrženo jeho využití panem Tai Dinh (Kyoto College of Graduate Studies for Informatics – Data Science and AI). Algoritmus negFIN se od ostatních algoritmů pro dolování frekventovaných sad liší tím, že dokáže zpracovat negativní množiny položek, což jsou množiny položek, které se v daném souboru dat často společně nevyskytují. Díky tomu je tento algoritmus užitečný pro aplikace, jako je detekce anomálií, kde nepřítomnost určitých položek může svědčit o anomálním chování. Celkově lze říci, že algoritmus negFIN je rychlý a efektivní algoritmus pro dolování častých množin položek, který dokáže zpracovat jak pozitivní, tak negativní asociace mezi položkami. Blíže je algoritmus představen v práci (Aryabarzan et al. 2018). Autoři v rámci článku algoritmus představili a srovnali jej s nejpoužívanějšími algoritmy pro dolování frekventovaných sad.

2.3 Frekventované sady

Při práci s technickými a odbornými texty je zásadní, aby byla použita sjednocená terminologie, která bude zahrnovat přesně definované pojmy a koncepty. V češtině není ustálený překlad termínu **Frequent Itemset** a **Frequent Itemset Mining / Analysis**. Dalším často používaným výrazem je **Market Basket Analysis (MBA)**. Ten jako jednu z metod data miningu uvádí (Petr 2014b; Šarmanová 2012). Pojmy **frekventované sady** a **analýza nákupního koše** lze chápat jako synonyma (Salmenkivi 2017). Někteří autoři se přiklánějí k použití termínu **frekventované sady instancí** (Dobešová 2022). K dosažení větší srozumitelnosti budou všechny tyto pojmy brány jako synonyma a v textu bude dále používán termín **frekventované sady**.

Frekventované sady lze definovat jako soubory položek, které se často vyskytují společně v datasetu s frekvencí překračující předem stanovený práh (threshold). Frekventované sady jsou klíčovým konceptem v dolování asociačních pravidel, které zahrnuje hledání vzorců a vztahů v datech (Agrawal a Srikant 1994). "Soubor položek je označen jako frekventovaný, pokud splňuje minimální hodnotu **prahu** (Minimum Threshold) pro hodnotu **podpory** (Support Count)" a že frekventované sady mohou být použity k identifikaci asociačních pravidel mezi položkami. (Agrawal a Srikant 1994). Podpora souboru položek je definována jako poměr počtu transakcí, ve kterých se frekventovaná sada vyskytuje vůči celkovému počtu transakcí. Je tedy vyjádřena v procentech. Podpora může být vyčíslena i jen jako absolutní počet transakcí, ve kterých se frekventovaná sada vyskytuje. Pro srovnání více sad transakcí (více nákupních košíků) je lepší procentuální vyjádření podpory.

V praxi mohou být frekventované sady použity k **identifikaci běžných souborů položek, které jsou často nakupovány spolu** zákazníky, jako jsou chléb, mléko a vejce v supermarketu. Tyto informace mohou být poté využity obchodníky pro optimalizaci umístění produktů, ceny a propagačních aktivit. Jedná se o tzv. MBA analýzu (ang. Market Basket Analysis) neboli analýzu nákupního koše. Představme si, že supermarket chce identifikovat podmnožiny produktů, které jsou často spolu kupovány. K tomuto účelu ukládá každou transakci zákazníka do databáze, kde řádky představují transakce a sloupce produkty (Tabulka 1). Hodnoty v tabulce ukazují, zda byl konkrétní produkt zakoupen v dané transakci. Cílem je najít frekventované sady, tedy množiny produktů, které se často vyskytují v téže transakci.

Tabulka 1 Vstupní tabulka pro příklad MBA, převzato z (Petr 2014b).

Id	Položky
1	A, B, E
2	B, D
3	B, C
4	A, B, D
5	A, C
6	A, B, C, E
7	A, B, C

Položky: A: mléko
B: chléb
C: cereálie
D: cukr
E: vejce

Dle (Petr 2014b) je vhodné zpracovat data do podoby transakční tabulky (Tabulka 2). Sloupce zastupují produkty a řádky reprezentují transakce. Každá buňka tabulky má hodnotu 1, pokud byl příslušný produkt zakoupen v dané transakci, a 0, pokud nebyl zakoupen. Úkolem je nalézt sady produktů, které se často vyskytují v jedné transakci (tj.

produkty, které jsou často zakoupeny společně) (Salmenkivi 2017). Metoda nepracuje s množstvím nakoupených věcí, a tedy zapisuje vždy každou položku pouze jednou.

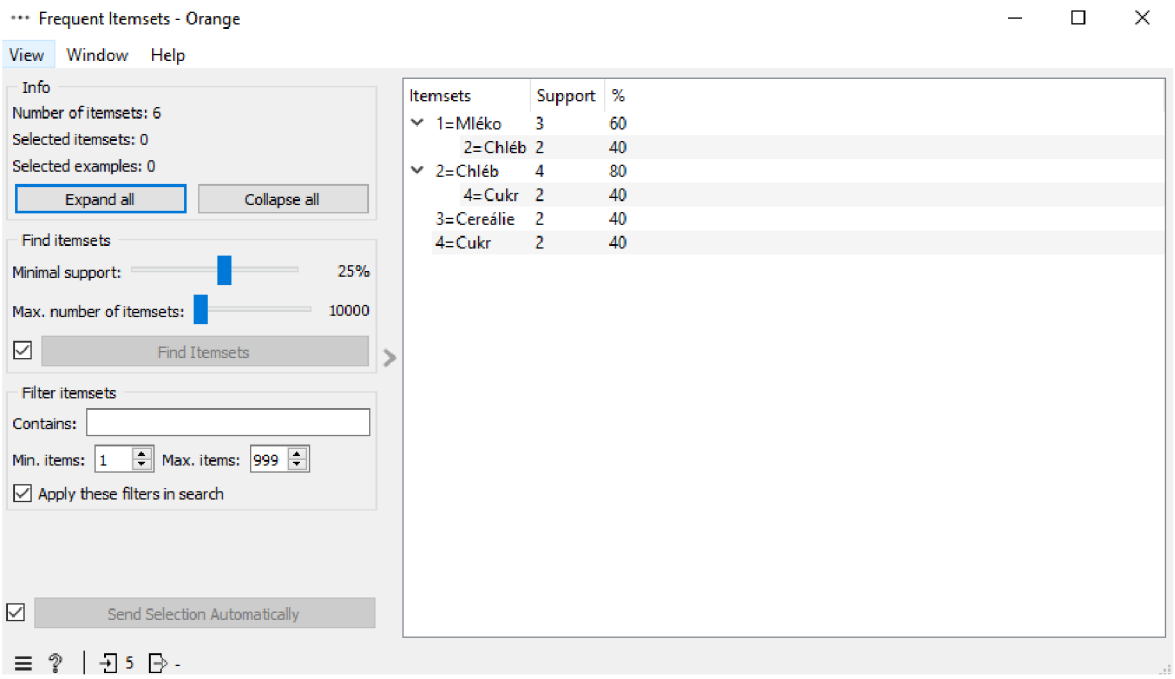
V MBA se **podpora (support)** používá jako míra frekvence výskytu konkrétních kombinací produktů v transakčním souboru dat. Podpora je definována jako pravděpodobnost, že daná sada (Itemset) se vyskytne v transakci, tedy jako poměr počtu transakcí obsahujících daný itemset a celkového počtu transakcí v souboru dat. Jak uvádí S. Han a M. Kamber v knize "**Data Mining: Concepts and Techniques**", podpora je důležitým ukazatelem v analýze nákupního košíku, protože umožňuje identifikovat významné kombinace produktů, které jsou nejčastěji kupovány společně a mohou být využity pro účely Cross-sellingu, tedy nabízení dalších produktů při nákupu. Vyšší hodnota podpory pro daný itemset znamená, že kombinace produktů je populární a je vhodná pro cross-sellingové kampaně (Han a Kamber 2012).

Tabulka 2 Transakční tabulka pro příklad MBA, převzato z (Petr 2014b).

Id	Mléko	Chléb	Cereálie	Cukr	Vejce
1	1	1	0	0	1
2	0	1	0	1	0
3	0	1	1	0	0
4	1	1	0	1	0
5	1	0	1	0	0

V následujícím odstavci je vysvětlena **podstata výpočtu frekventovaných sad** s interpretací výsledků na příkladu transakční tabulky (Tabulka 2). Generování bylo provedeno v programu Orange. Uživatel definuje minimální podporu v procentech (posunem modrého obdélníku na stupnici) (Obrázek 2). V příkladu níže je aplikována minimální podpora 25 %. Podpora 25 % znamená, že aby byla sada frekventovaná, musí se nacházet alespoň v jedné čtvrtině všech vstupních transakcí. Vygenerované sady v programu Orange mají hierarchickou strukturu. Sady začínají na jednoprvkových a rozvětvují se do víceprvkových (Obrázek 2). Frekventovaná sada je zapsána takto Itemset = Mléko, Support = 3, %= 60. To značí, že mléko samotné se nachází ve třech transakcích. Jedná se tedy o jednoprvkovou frekventovanou sadu s podporou 60 %, která přesahuje definovaný 25% práh. Obdobně je iterováno přes všechny jednoprvkové kandidátní sady a pokud překročí definovanou minimální podporu jsou zapsány do výsledků. V tomto případě jdou všechny jednoprvkové sady až na vejce nad stanovaným prahem. V případě vajec se nacházejí v 1 z 5 transakcí, tedy mají podporu 20 %. Je důležité říci, že **pokud není jednoprvková sada frekventovanou, nemůže vstoupit do žádné dvou či víceprvkové sady**. Dále platí, že čím je frekventovaná sada víceprvková, tím její podpora klesá. Můžeme tedy říci, že vejce se nebudou nacházet v žádné dvouprvkové sadě. Generování pokračuje hledáním dvouprvkových sad. Ty jsou ve vizualizaci výsledků v programu Orange znázorněny odsazením řádku pod jednoprvkovou sadou. V rámci datasetu existují dvě dvouprvkové sady a to Mléko, Chléb s podporou 40 % a Chléb, Cukr rovněž s podporou 40 %. V rámci datasetu neexistují žádné tří či víceprvkové frekventované sady. Celkem bylo vypočítáno 6 frekventovaných sad.

Na příkladu 20 domů s atributy vzdálenosti od vody, vzdálenosti od hlučné silnice, tvaru reliéfu a ceny vysvětluje v knize Orange, Praktický návod do cvičení předmětu Data Mining (Dobešová 2022) aplikaci metody frekventovaných sad. Princip fungování metody je blíže popsán také v knize Metody Data Miningu – část 2 (Petr 2014b).



Obrázek 2 Nástroj Frequent Itemsets v programu Orange – minimální podpora 25 %, vstupní data (Tabulka 2).

Primárním cílem analýzy frekventovaných sad je odhalit vztahy mezi různými položkami v datasetu a identifikovat ty, které se vyskytují nejčastěji. Tuto metodu lze použít v různých oblastech, jako jsou například:

- **Obchodní analýza**: k identifikaci produktů, které se nejčastěji kupují společně a následně pro plánování strategie prodeje a marketingu.
- **Zdravotnická a biologická věda**: k identifikaci souvislostí mezi symptomy, léky a chorobami, což může pomoci v diagnostice a léčbě.
- **Finanční analýza**: k identifikaci vztahů mezi investicemi a akciemi na burze.

V podstatě se analýzy frekventovaných sad používají tam, kde je potřeba odhalit vztahy mezi položkami v datasetu a identifikovat nejčastější kombinace. Většina aplikací analýzy je zaměřena na neprostorová data. Diplomová práce se zaměřuje na aplikační stránku, která doposud nebyla dominantní, a to **aplikaci v oblasti prostorových dat**.

2.4 Asociační pravidla

Asociační pravidla jsou technikou dolování dat, která se používá k identifikaci vzorů a vztahů mezi různými proměnnými v souboru dat. Tato pravidla se používají k odhalení základních asociací nebo korelací mezi dvěma nebo více proměnnými a lze je použít v široké škále oblastí. Asociační pravidla jsou obvykle reprezentována ve formě zápisu "if-else", kde antecedent neboli levá strana pravidla představuje podmínku nebo vstupní proměnnou a konsekvent neboli pravá strana představuje výslednou nebo výstupní proměnnou (Han a Kamber 2012). Síla asociačního pravidla se obvykle měří pomocí metrik, jako je support, confidence a lift, které udávají četnost, přesnost a významnost pravidla. Jednou z klíčových aplikací asociačních pravidel je analýza tržního koše, kde se používají k určení, které produkty zákazníci často nakupují společně. Asociační pravidla lze také použít k identifikaci potenciálních příležitostí ke křížovému prodeji, ke zlepšení segmentace zákazníků a k optimalizaci obchodních procesů (Berry a Linoff 2004).

2.5 Zahraniční a domácí studie

Understanding Spatial Concentrations of Road Accidents Using Frequent Item Sets (Geurts et al. 2005) – Pochopení prostorové koncentrace dopravních nehod pomocí analýzy frekventovaných sad

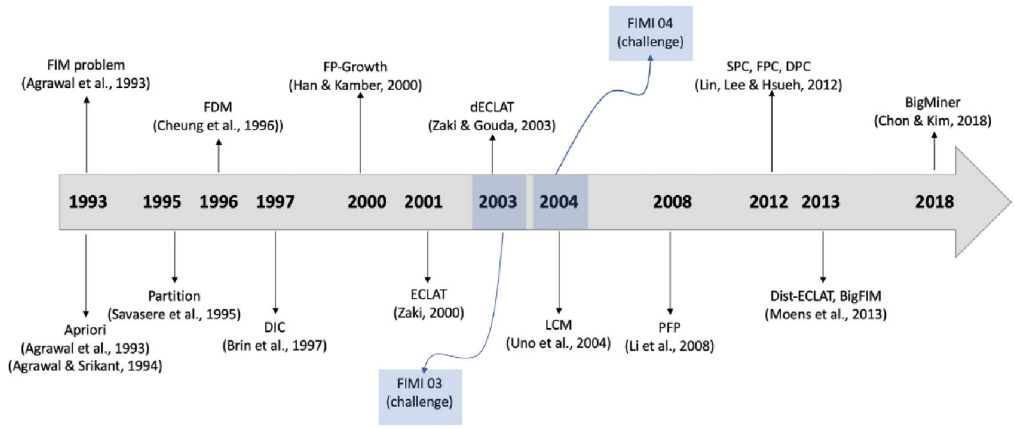
Jednou z oblastí, ve které byla aplikována analýza frekventovaných sad, je studium dopravních nehod. Ve svém článku zkoumá Geurts et. al., proč dochází k dopravním nehodám na určitých úsecích silnic, a používá techniku frekventovaných sad k identifikaci společných okolností nehod. Studie se zaměřuje na porovnání charakteristik nehod v "černých" zónách (nebezpečné lokality) oproti rozptýleným místům v belgickém přímořském regionu. Výsledky ukazují, že nehody v "černých" zónách zahrnují odbočování vlevo na signalizovaných křižovatkách, srážky s chodci, ztrátu kontroly nad vozidlem a deštivé počasí, zatímco nehody mimo "černé" zóny zahrnují odbočování vlevo na křižovatkách s dopravním značením, čelní srážky a opilé účastníky silničního provozu. Studie naznačuje, že neexistuje jediné řešení pro snížení nehodovosti (Geurts et al. 2005).

A Survey of Itemset Mining (Fournier- Viger et al. 2017) – Přehled vytěžování datových sad

Článek se zabývá oblastí vytěžování frekventovaných sad a jeho různými aplikacemi. Pojednává také o rozšířeních základního problému, aby bylo možné pracovat s dynamickými databázemi, neurčitými daty a omezeními. Článek se také dotýká příbuzných problémů dolování vzorů, včetně dolování vzorů, dolování asociačních pravidel. Nakonec článek upozorňuje na možnosti výzkumu a dostupné open-source implementace algoritmů pro dolování vzorů (Fournier- Viger et al. 2017). V rámci spolupráce s panem Tai Dinh byl rovněž doporučen program SPMF autora tohoto článku (Fournier-Viger 2023) Jedná se o open-source data miningovou knihovnu napsanou v jazyce Java, která implementuje 254 data miningových algoritmů včetně v práci využívaného algoritmu negFIN. Podrobněji je program popsán v podkapitole SPMF (Sequential Pattern Mining Framework).

Frequent Itemset Mining: A 25 Years Review (Luna et al. 2019) – Vytěžování frekventovaných sad: přehled za 25 let

Článek autorů (Luna et al. 2019) se zabývá přehledem existujících algoritmů pro FIM (Frequent Itemset Mining). Dělí algoritmy a jejich řešení na sekvenční, vícevláknová a s distribuovaným výpočtem. Cílem tohoto článku je ukázat zlepšení, která byla provedena za posledních 25 let (Obrázek 3), tj. od doby, kdy byla úloha FIM poprvé popsána (Agrawal a Srikant 1994).



Obrázek 3 Časová osa výzkumu frekventovaných sad během 25 let, převzato z (Luna et al. 2019).

From Frequent Itemsets to Semantically Meaningful Visual Patterns (Yuan et al. 2007) – Od frekventovaných sad k sémanticky významným vizuálním vzorům

Ve své práci se Yuan et. al. věnovali objevování smysluplných vizuálních vzorů v obrazových databázích, které se od textových a transakčních dat liší svými vysokodimenzionálními vlastnostmi a prostorovou strukturou. Autoři navrhli nový přístup, který kombinuje techniky dolování frekventovaných sad, samonaváděného shlukování a sumarizace vzorů, aby se s těmito obtížemi vypořádali. Jejich metoda dokáže účinně a efektivně objevovat sémanticky významné vzory v reálných obrazech, jak dokazují experimentální výsledky. Tento přístup aplikovali na reálné obrazy automobilů a obličejů a úspěšně objevili sémanticky významné vizuální vzory (Yuan et al. 2007).

Obdobnou možnost aplikace nastiňuje ve své knize věnující se programu Orange pro výuku data miningu Dobešová. V rámci 8. kapitoly uvádí problematiku neuronových sítí a doplňku Image Analytics a v podkapitole 8.2 jej aplikuje k vyhledávání vizuálně podobných mapových výřezů (Dobešová 2022).

The Similarity of European Cities Based on Image Analysis (Dobesova 2019) – Podobnost evropských měst na základě analýzy obrazu

Článek se zaměřuje na možnost aplikace neuronových sítí a klasifikaci obrazu ke zkoumání podobnosti evropských měst. Autorka využívá data Urban Atlas v rámci projektu Copernicus. K analýze dat byla použita natrénovaná neuronová síť Painters, Embeder doplňku Image Analytics v rámci programu Orange. Tvůrcem neuronové sítě je Nejč Ilenič (Kaggle 2016). Dále bylo použito hierarchické shlukování k seskupení měst s podobnými vzorci městské struktury, zeleně nebo tvarů ploch. Analyzováno bylo 100 evropských měst o velikosti 50 až 200 tis. obyvatel. V článku je uveden seznam nejpodobnějších dvojic měst na základě těchto vzorů. Je zmíněno, že je uvažováno pouze rozložení a druh využití území, nikoliv počet obyvatel nebo velikost jednotlivých území. Výsledky jsou slibné a vhodné k dalšímu výzkumu. Dále je nastíněna možnost aplikace použitých postupů pro výuku předmětu Data Mining, kterého je autorka garantem v rámci magisterského studia. Možnostem aplikace dat Evropské Unie ve výuce geoinformatických předmětů se věnuje (Dobešová et al. 2022).

Experiment in Finding Look-Alike European Cities Using Urban Atlas Data (Dobesova 2020) – Experiment k nalezení podobných evropských měst s užitím dat Urban Atlasu

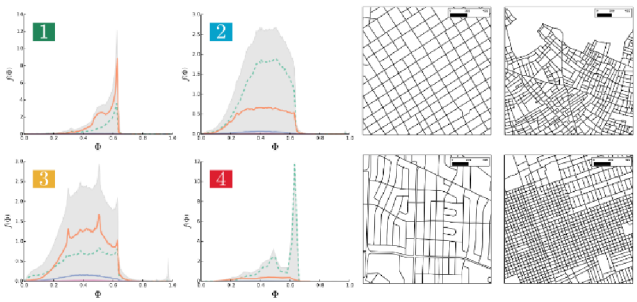
(Dobesova 2020) pojednává o experimentu, který využívá strojové učení k identifikaci měst, která jsou si podobná na základě údajů o využití území (landuse). Studie použila datovou sadu Copernicus European Urban Atlas 2012 a předem natrénovanou neuronovou síť Painters (software Orange) k identifikaci podobných měst na základě vzorců využití půdy. Analyzováno bylo přibližně 800 evropských měst. Výsledkem výzkumu bylo odhalení vzorců a podobností mezi městy na základě jejich uspořádání využití území. Příklady podobných měst byly identifikovány na základě jejich dominantních kategorií využití území, uspořádání komerčních a průmyslových ploch, okolního přírodního prostředí a dalších faktorů. K vyhledání byla použita metoda k-Nearest Neigbour nad feature vektorem získaným z neuronové sítě Painters.

Discovery of Spatial Association Rules in Geographic Information Databases (Koperski a Han 1995) – Nalezení prostorových asociačních pravidel v geodatabázích

Koperski a Han (1995) ve své práci představují rozšíření dolování asociačních pravidel na prostorová data. Studie probíhala na městech kanadské provincie Britská Kolumbie a mezi zkoumané parametry měst patřila blízkost k vodě, komunikacím, administrativním hranicím či dolům.

A Typology of Street Patterns (Louf a Barthelemy 2014) – Typologie vzorů uliční sítě

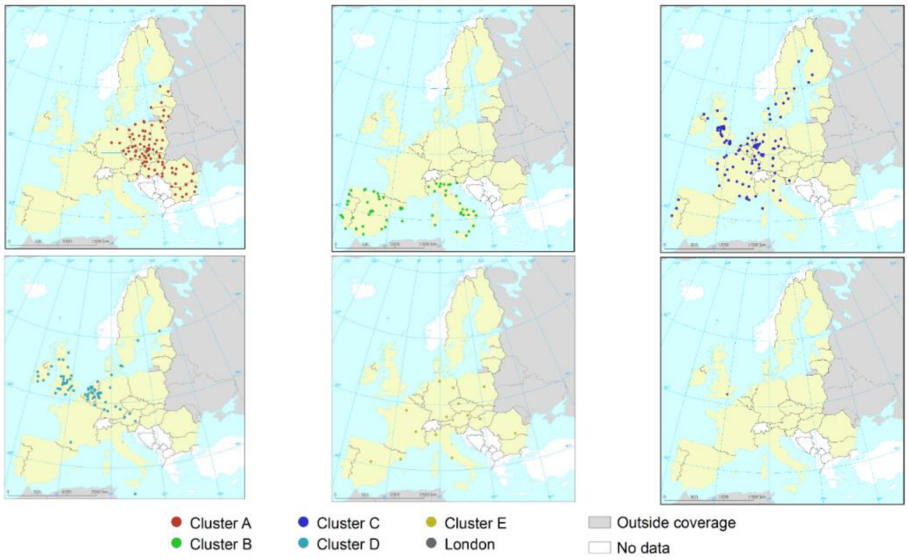
Autoři (Louf a Barthelemy 2014) navrhují metodu klasifikace měst na základě jejich uličních vzorů pomocí analýzy podmíněného pravděpodobnostního rozdělení faktoru tvaru bloků s danou rozlohou. Dochází k závěru, že metodou lze vytvořit "otisk prstu" města, který lze následně použít jako základ pro typologii měst pomocí metody hierarchického shlukování. Autoři tuto metodu aplikují na 131 měst a nacházejí 4 velké rodiny měst (Obrázek 4), které se vyznačují různou četností bloků o určité ploše a tvaru. Ukazují také, že otisk města lze považovat za součet otisků charakterizujících různé čtvrti uvnitř města. Metoda může pomoci pochopit příčiny, které stojí za odlišnými tvary městských ulic. Na obdobné téma publikoval Boeing několik studií (Boeing 2018; 2019; 2020).



Obrázek 4 Čtyři skupiny měst podle struktury uliční sítě, převzato z (Louf a Barthelemy 2014).

Similarities and Diversity of European Cities. A typology Tool to Support Urban Sustainability (Mirko et al. 2018) – Podobnosti a rozdílnosti evropských měst: typologický nástroj pro urbánní udržitelnost

Zpráva autorského kolektivu (Mirko et al. 2018) Evropské enviromentální agentury (European Environment Agency – European Topic Center on Urban Land and Soil Systems) a několika spolupracujících univerzit představuje **typologický nástroj založený na analýze údajů o využití území** z téměř 385 evropských měst, který podporuje udržitelnost měst. Nástroj rozděluje města do šesti typů na základě jejich vzorců a charakteristik využití území, včetně městských jader, polycentrických, příměstských, průmyslových, lesních a zemědělských měst. Autoři tvrdí, že tento nástroj může pomoci tvůrcům politik při pochopení rozmanitosti a společných rysů evropských měst a při vytváření politik na míru na podporu udržitelného rozvoje měst. Každý ze šesti shluků (Obrázek 5) je charakterizován socioekonomickými a fyzickogeografickými charakteristikami. Jsou uvedena nejreprezentovanější města daného shluku s charakteristikou.



Obrázek 5 Prostorová distribuce shluků měst na základě typologie, převzato z (Mirko et al. 2018).

2.6 Diplomové a bakalářské práce

Porovnání urbánního prostoru pomocí kruhových výsečí (Janoušek 2019) – Comparison of Urban Area by Circular Sectors

Ve své diplomové práci se (Janoušek 2019) věnoval porovnávání urbánních ploch vybraných měst. Navázal na bakalářskou práci, v rámci které vznikl nástroj Area Index From Circular Sectors (Janoušek 2017). Výsledky se skládaly z vypočítaných plošných indexů v kruhových výsečích, korelačních a shlukových analýz dat řady měst a vizualizace výsledků. Byla nalezena podobná města na základě hodnot indexů v kruhových výsečích ze zdrojových dat Urban Atlas z projektu Copernicus. V rámci diskuse autor naznačuje, že způsob umístění počátečních bodů při výpočtu plošných indexů výrazně ovlivňuje konečné výsledky. Dalším z parametrů metody, která ovlivňuje výsledky je počet segmentů a kružnic. Pro porovnání výsledků je tedy důležité zachovat jednotnou šablonu pro všechna města (Dobesova 2019).

Podobnost evropských měst a jejich funkčních území (Urbančík 2022) – The Similarity of European Cities and Their Functional Areas

Cílem diplomové práce (Urbančík 2022) bylo aplikovat různé metriky podobnosti a nepodobnosti na parametry využití území (landuse) Evropských měst. V rámci práce byly vypočítány krajinné metriky, a to jak nástrojem FRAGSTATS, tak nástrojem Patch Analyst. Následně byla provedena hierarchická shlukovací analýza pro nalezení podobných měst. Celkově bylo analyzováno 100 evropských měst.

Aplikace vyhledávání kolokačních vzorů na prostorová data (Bučková 2022) – Application of Co-location Patterns Searching on Spatial Data

Autorka (Bučková 2022) se ve své diplomové práci věnovala analýze kolokačních vzorům v prostorových datech. Podrobně zkoumala nástroj Colocation Analysis v ArcGIS Pro, se zaměřením na bodová data. Nástroj aplikovala ve třech případových studiích, přičemž v jedné z nich aplikovaný postup použila k tvorbě manuálu. První případová studie se zaměřovala na zdravotnická zařízení a lékárny a jejich kolokaci. Autorka se blíže seznamovala s nástrojem a blíže zkoumala podstatu zkoumaného jevu. Ve druhé studii byla zkoumána data kriminality na území Filadelfie, USA. Specifikem této studie bylo přidání časového aspektu do vyhledávání kolokačních vzorů. Poslední studie se věnovala námořnímu pirátství. Analýza kolokací v ArcGIS se vztahuje na dvě předem definované třídy prvků, například zdravotnická zařízení a lékárny. Nelze vyhodnotit kolokace mezi více než dvěma třídami prvků. Tento postup umožňuje identifikovat frekvenci sousedících prvků prostřednictvím počtu dvojic kolokovaných prvků. Jedná se tedy o podobnou analýzu jako analýza frekventovaných sad, ale s omezením na jeden vstupní soubor prvků, ke kterému se vyhodnocují kolokace s druhým souborem prvků. Analýza kolokací představuje zjednodušený případ vyhodnocení současného výskytu prvků do určité vzdálenosti.

Aplikace asociačních pravidel na prostorová data (Trnová 2020) – Application of Association Rules on Spatial Data

Práce se věnovala generování asociačních pravidel na vhodných prostorových datech. Autorka (Trnová 2020) vytvořila 3 modely, které lze aplikovat na prostorová data, která jsou po aplikovaných úpravách použitelná ke generování asociačních pravidel. V rámci diskuse autorka uvádí, že přínosem práce je samotná schopnost generování asociačních pravidel bez zkoumání jejich přínosnosti. Důležitým výstupem práce je podrobný návod pro zpracování prostorových dat.

2.7 Přínos pro diplomovou práci

Tato práce se zabývá **aplikací data miningových postupů v geoinformatickém kontextu** a **zkoumá možnosti využití dolování frekventovaných sad na prostorových datech**. V posledních letech se zájem o aplikaci data miningu v geoinformatických aplikacích zvyšuje, protože může poskytnout nové poznatky. Zatímco prostorová data mají v rámci celého Data Miningu jen okrajové postavení, geoinformatika může přispět k vylepšení interpretace výsledků. Jeden z hlavních přínosů aplikace data miningu v prostorových datech je schopnost detekovat vzorce a souvislosti, které by jinak zůstaly skryty v množství dat. Výsledky získané z dolování frekventovaných sad mohou být využity v mnoha různých oblastech geoinformatiky a urbanismu, jako například **při plánování měst, dopravní infrastruktury, při sledování stavu a vývoje městských oblastí nebo při analýze vlivu změn klimatu na životní prostředí.**

V rámci rešerše byl zkoumán **informatický pohled na problematiku**, který se převážně věnuje tématům jako přehled použitelných algoritmů, jejich složitost a výpočetní nároky. Vývoj algoritmů v oblasti frekventovaných sad zaznamenal za posledních 20 let významný pokrok. Zlepšil je jak výpočetní výkon počítačů, tak i velikost vstupních dat (Big Data). V rámci rešerše byly prozkoumány jednotlivé algoritmy a na základě konzultace s odborníkem na data mining byl zvolen jeden z nejnovějších algoritmů negFIN.

V rámci odborné rešerše byly zkoumány jednotlivé aplikace geografických problematik s použitím Data Mining. Větší důraz byl kladen na socioekonomickou sféru geografie z důvodu zaměření práce. Velmi přínosnou byla práce autorského kolektivu (Mirko et al. 2018) Evropské enviromentální agentury, která představila **typologický nástroj založený na analýze údajů o využití území** z téměř 385 evropských měst. Práce zkoumala velké množství faktorů (počet obyvatel, hustotu zalidnění, nárůst a pokles počtu obyvatel, věkovou strukturu, turismus, kvalitu ovzduší, ohroženost chudobou a další). Celkově jako nejpřínosnější byly práce věnující se aplikaci evropských dat (Dobesova 2019; 2020; Mirko et al. 2018; Janoušek 2019; Urbančík 2022).

3 METODY A POSTUP ZPRACOVÁNÍ

V oblasti zkoumání podobnosti měst existuje mnoho přístupů. V rámci diplomové práce je blíže zkoumána možnost aplikace techniky **frekventovaných sad**, která je klasickou metodou data miningu (Salmenkivi 2017). Tato technika je aplikována na data využití území (land use), která jsou dostupná na celoevropské úrovni prostřednictvím projektu Copernicus.

Projekt Copernicus je iniciativou Evropské unie zaměřenou na vývoj integrovaného systému pro správu a monitorování životního prostředí. Jeho cílem je poskytovat včasné, přesné a spolehlivé informace o životním prostředí, změně klimatu a přírodních katastrofách. Městská složka projektu se zaměřuje na poskytování podrobných informací o městském prostředí, například o způsobech využívání území, hustotě osídlení a kvalitě životního prostředí. Klíčovým výstupem projektu je **Urban Atlas**, který poskytuje **podrobné informace o využití území** pro více než 700 městských oblastí v Evropě.

3.1 Použité programy

3.1.1 ArcGIS Pro 2.8.3 a ArcGIS ModelBuilder

V rámci práce byl použit program **ArcGIS Pro**, jakožto stěžejní GIS program. Sloužil převážně k přípravě prostorových dat a k jejich následnému zpracování. **ArcGIS Model Builder** byl použit k implementaci nástroje pro přípravu dichotomických dat, který byl použit k následnému generování vstupních dat pro generování frekventovaných sad sousednosti polygonů využití území.

ArcGIS Pro ModelBuilder je nástroj grafického uživatelského rozhraní (GUI), který slouží k vytváření, úpravám a správě pracovních postupů nebo modelů v softwaru ArcGIS Pro. Umožňuje uživatelům automatizovat a zefektivnit složité úlohy geoprostorové analýzy a zpracování dat, což usnadňuje vytváření opakovatelných a konzistentních výsledků. Pomocí nástroje ModelBuilder mohou uživatelé vytvářet a upravovat složité modely geoprocessingu přetažením nástrojů na plátno a jejich propojením se vstupními a výstupními parametry. Nástroj podporuje širokou škálu formátů prostorových dat a uživatelé mohou do pracovních postupů snadno integrovat skripty Pythonu a další vlastní nástroje. Celkově je ModelBuilder výkonný nástroj, který zvyšuje efektivitu a produktivitu geoprostorových analytiků, výzkumníků a dalších odborníků pracujících s prostorovými daty.

3.1.2 MS Excel

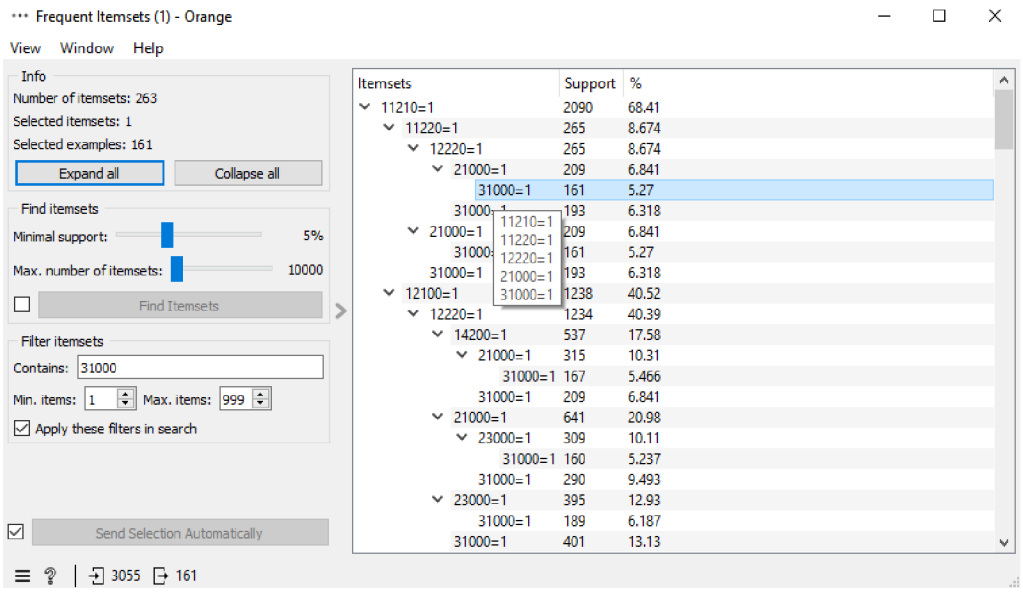
Výsledky, které generuje nástroj vytvořený v ArcGIS Pro ModelBuilder, jsou prezentovány v podobě Excelových tabulek. Tyto tabulky obsahují informace o sousednosti jednotlivých polygonů, kde hodnota 1 indikuje, že dané využití území se vyskytuje v sousednosti s daným polygonem a tyto hodnoty společně tvoří transakce. Výstup z tohoto nástroje lze tedy označit jako transakční tabulky. Pro tvorbu takovýchto tabulek byl využit program Microsoft Excel. Data v tomto formátu jsou dále použitelná v programu Orange. Program byl rovněž využit pro tvorbu tabulkových a grafových výstupů v rámci práce.

3.1.3 Orange 3.34.0

Orange je open-source software pro dolování dat, který uživatelům poskytuje vizuální programovací prostředí pro analýzu dat a strojové učení. Vyvinula jej Laboratoř bioinformatiky na Univerzitě v Lublani ve Slovinsku a je k dispozici zdarma uživatelům.

Software je navržen tak, aby uživatelům pomáhal snadno manipulovat, vizualizovat a analyzovat rozsáhlé soubory dat, aniž by vyžadoval jakékoli znalosti programování.

Zásuvný modul "Associate" v aplikaci Orange je výkonný nástroj pro dolování frekventovaných sad a učení asociačních pravidel. Používá algoritmus FP-growth. Orange umožňuje uživatelům získávat z dat časté vzory a vztahy a počítá pro nalezené frekventované sady podporu (Support) jak v absolutních hodnotách, tak v procentech. Velkou nevýhodou je nemožnost exportu výsledků do strojově čitelné podoby, a tedy nutného ruční zpracování. Z tohoto důvodu bylo od programu Orange po testování upuštěno. Určitou výhodou je, že program Orange umožňuje filtrovat frekventované sady pomocí řádku Contains. Při vyplnění hodnoty 31000 (lesy) jsou vyfiltrovány pouze sady obsahující tuto hodnotu. Zvýrazněna je pětiprvková frekventovaná sada obsahující hodnoty 11210, 11220, 12220, 21000 a 31000 značící nesouvislou vysoce hustou městskou zástavbu, nesouvislou středně hustou městskou zástavbu, ostatní silnice, ornou půdu a lesy s podporou 5,27 % (Obrázek 6). Jak číst výsledky bylo vysvětleno v kapitole 2.3 Frekventované sady.



Obrázek 6 Rozhraní programu Orange – nástroj Frequent Itemsets.

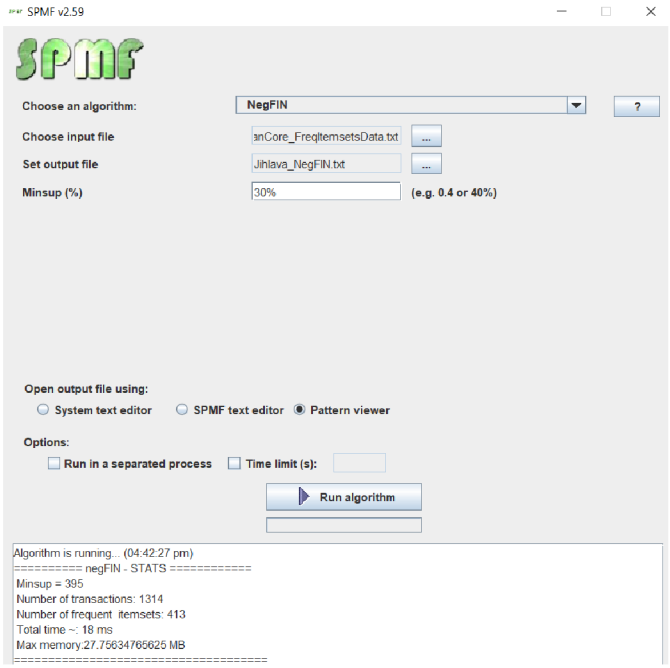
3.1.4 SPMF (Sequential Pattern Mining Framework)

Program SPMF je **open source data miningová knihovna**, kterou vyvinul Philippe Fournier-Viger a jeho tým. Bližší popis knihovny je dostupný na internetových stránkách (https://www.philippe-fournier-viger.com/spmf). Současná verze je 2.59 a byla spuštěna 25. prosince 2022. Stránky obsahují mimo popisu knihovny a návodu pro stažení také detailní popis jednotlivých implementovaných algoritmů.

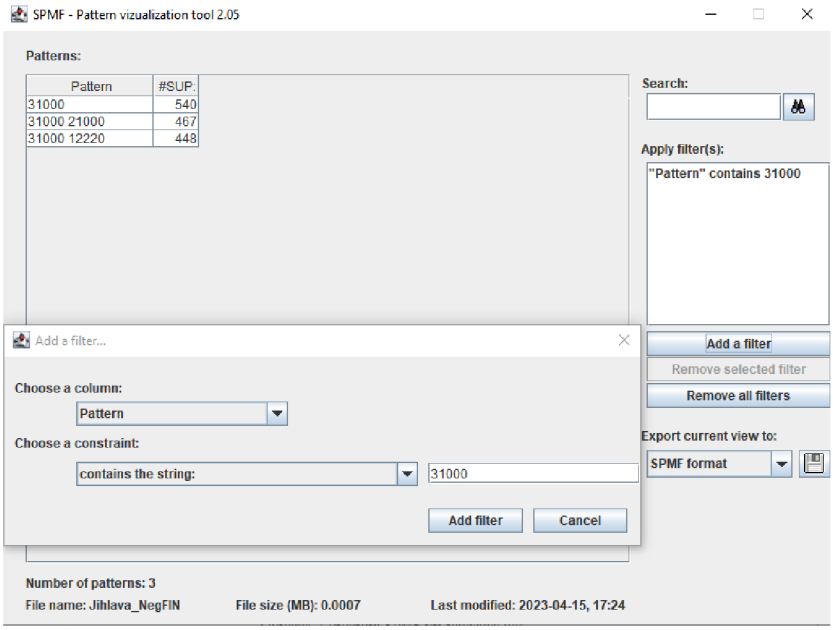
V sekci **Algoritmy** se nachází přehled všech implementovaných algoritmů s odkazy na odbornou literaturu (první uvedení jednotlivých algoritmů). V sekci **Dokumentace** se nachází již zmíněný detailní popis včetně typových úloh doplněný o příkladová data. V jednotlivých dokumentacích jsou popsána vstupní data, výstupní data, příklad správné interpretace výsledků, struktura vstupních a výstupních souborů a odkazy na další informace. V sekci **Videa** jsou obsaženy přednášky k jednotlivým algoritmům včetně praktických ukázek. Celkově lze program považovat za vhodný pro výuku data miningu, popřípadě pro samostudium. Zde jsou velkým přínosem právě video přednášky.

Ve verzi s grafickým rozhraním je obsaženo 230 algoritmů. Verze ve zdrojovém kódu obsahuje 254 algoritmů. **Hlavním kladem aplikace** je její grafické rozhraní, široké spektrum implementovaných algoritmů, a především rozhraní **Pattern Viewer**, které umožňuje filtrovat napočítané frekventované sady v našem případě na základě jednotlivých kódů využití území. Můžeme tedy vyhledat všechny frekventované sady, které obsahují například kód využití území 13000. Výsledky programu lze exportovat do textového souboru. Nevýhodou je, že nepočítá relativní hodnoty podpory v procentech.

Uživatel nahrává vstupní soubor ve formátu txt, nastavuje umístění pro uložení, specifikuje další parametry. V případě algoritmů při hledání frekventovaných sad uživatel zadá minimální podporu (Obrázek 7).



Obrázek 7 Program SPMF při volání algoritmu negFIN na datech města Jihlava.



Obrázek 8 Program SPMF při realizaci algoritmu negFIN – Pattern Viewer.

3.1.5 FI.ipynb a negFI.py

V rámci konzultací s expertem na data mining panem Tai Dinh z Univerzity Kyoto v Japonsku byl pro potřeby generování frekventovaných sad poskytnut Jupyter Notebook FI.ipynb a Python programový kód funkce negFI.py. Jupyter notebook je webové interaktivní výpočetní prostředí, které lze použít k vytváření dokumentů obsahujících jak počítačový kód (např. Python), tak jiné textové prvky. Notebook FI.ipynb volá funkci negFIN a hromadně zpracovává všechna vstupní data nalezená ve vstupním adresáři „input_datasets". Automaticky ukládá textové soubory s frekventovanými sadami a jejich podporou (v procentech i absolutní hodnotě) pro každý ze vstupních souborů. (příloha: Vstupni_Data/JupyterNotebook_FI/FI.ipynb).

3.2 Použitá data

V rámci práce je vybráno 100 evropských měst. Jako podklad byla využita příloha diplomové práce Janouška (příloha: Vstupni_Data/MSExcel_soubory/cities100_Janousek .xlsx). Výběr probíhal obdobně jako v diplomové práci (Janoušek 2019), kdy byla vybrána města v rozsahu počtu obyvatel 50 000-200 000 obyvatel. Tento rozptyl je v datové sadě nejčetnější a obsahuje města z většiny států. Vstupní data Urban Atlsu neobsahují údaj o počtu obyvatel a z tohoto důvodu byl převzat ze statistik Eurostatu. Důvodem k omezení velikosti měst byla následná potřeba srovnatelnosti výsledků. Blíže se metodě výběru měst věnuje kap. 4.1 Výběr dat.

Menší část (22) tvoří města ze studie **Experiment k nalezení podobných evropských měst s použitím data Urban Atlasu** (Dobesova 2020), který je blíže představen v kapitole 2.5 Zahraniční a domácí studie. Výsledkem výzkumu bylo odhalení vzorců a podobností mezi městy na základě jejich uspořádání využití území. Výsledky tvoří vždy dvojice podobných měst. Tato data budou využita k porovnání získané podobnosti v rámci studie a aplikace metody frekventovaných sad v případové studii 5.4 Porovnání s výsledky studie (Dobesova 2020).

3.3 Postup zpracování

Prvním krokem byla rešerše odborné literatury a konzultace s odborníkem na data mining p. Tai Dinh. Následně byl navržen postup zpracování (Obrázek 10).

Před samotnou analýzou dat a data miningem bylo potřeba **vytvořit nástroj** v rámci ArcGIS Pro, který by generoval data vhodná pro použití v data miningu. Celkem byly navrženy 3 nástroje, ze kterých po testování výkonosti a správnosti výsledků byl vybrán nástroj **SearchDistinctLanduse_SpatialJoin** blíže představený v kapitole 4.3 Implementace nástroje pro přípravu kategoriálních a dichotomických dat. Tento nástroj umožňuje převádět prostorová data, do formátu, který je kompatibilní s programem Orange, SPMF a Jupyter Notebookem FI.ipynb. Výstupem nástroje jsou tabulková **transakční data**, jejichž struktura je blíže představena v kapitole 4.3.3 SearchDistinctLanduse_SpatialJoin. Celkově vytvoření tohoto nástroje výrazně usnadňuje práci s prostorovými daty a zvyšuje efektivitu jejich analýzy a umožňuje aplikaci metody frekventovaných sad. Původně zamýšleným programem pro samotný data mining byl program Orange. Ten se ale ukázal jako neaplikovatelný z důvodu nemožnosti exportovat výsledky do strojově čitelné podoby. Data mining probíhal v prostředí programu **SPMF a Jupyter Notebooku FI.ipynb**.

Následně byla data vybraných měst stáhnuta ze stránek Urban Atlas (https://land.copernicus.eu/local/urban-atlas/urban-atlas-2018). Pro každé město byla vyříznuta pouze oblast urbánního jádra (urban core) pro rok 2018. K tomu byl

implementován **script v jazyce Python – Batch Clip (Urban Atlas)**. Následně byla pomocí realizovaného nástroje vypočítána transakční data sousednosti pro další analýzu. Zadání práce specifikuje vytvoření kategoriálních nebo dichotomických dat, z tohoto důvodu je výstupem programu jednak textový soubor a také soubor MS Excel.

Frekventované sady pro každé město byly vypočítány pomocí programu FI.ipynb. Následně byly odstraněny duplicity pomocí scriptu Aggregation (4.6 Příprava dat pro případovou studii – evropská města). Frekventované sady byly graficky zpracovány do podoby tabulky (Obrázek 9) s podbarvením odpovídajícím legendě Urban Atlas s použitím scriptu TXTtoEXCEL.py a podmíněného formátování. Script převání vypočítané textové soubory programu FI.ipynb do výsledné Excelové tabulky. Jednotlivé listy souboru obsahují vybraných 100 měst s vypočítanými frekventovanými sadami sousednosti využití území s minimální podporou 5 %.

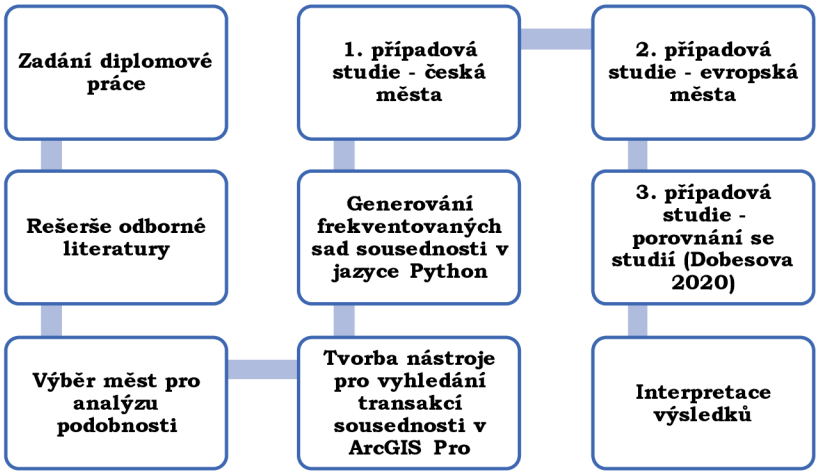
	A	B	C	D	E	F	G	H	I
1	% Podpo	Abs. Podpo	Landuse1	Landuse2	Landuse3	Landuse4	Landuse5	Landuse6	Landuse7
2	97	1266	12220						
3	73	948	12100						
4	72	941	12220	12100					
5	69	898	11210						
6	69	898	12220	11210					
7	61	802	11100						
8	61	802	12220	11100					
9	60	789	21000						
10	58	755	12220	21000					
11	51	671	12100	11100					
12	51	671	12220	12100	11100				
13	49	637	11210	11100					
14	49	637	12220	11210	11100				
15	49	634	12100	11210					
16	49	634	12220	12100	11210				
17	44	579	14100						
18	44	578	12220	14100					
19	41	541	14200						
20	41	538	12220	14200					
21	40	522	11210	21000					
22	40	522	12220	11210	21000				
23	40	524	12100	21000					
24	40	519	12220	12100	21000				
25	40	519	12100	11210	11100				
26	40	519	12220	12100	11210	11100			
27	39	515	11100	14100					
28	39	515	12220	11100	14100				
29	39	505	12100	14100					
30	39	504	12220	12100	14100				
31	35	463	12100	11100	14100				
32	35	463	12220	12100	11100	14100			
33	35	452	11210	14100					

PRESOV | ZILINA | CAMBRIDGE | LINCOLN | WORCESTER

Obrázek 9 Grafická vizualizace vypočítaných frekventovaných sad.

Prvním cílem bylo představení možné intepretace zjištěných frekventovaných sad, což bylo provedeno pro dvě vybraná města Cheltenham a Prešov. Byly identifikovány typické frekventované sady sousednosti doplněné o mapové výřezy a slovní popis. Nejedná se o podobná města, kapitola má za cíl představit možnou interpretaci výsledků frekventovaných sad pro jedno město samostatně.

V návazné části byly provedeny tři případové studie. První se zaměřovala na česká města a jejím cílem bylo popsání charakteru všech 15 českých měst v rámci datasetu Urban Atlas 2018. Cílem bylo komplexně popsat charakter českých měst jako celku. Druhá se zaměřovala na vybraných 100 evropských měst, z nich část tvořily dvojice podobných měst zjištěných v rámci studie (Dobesova 2020). Cílem třetí případové studie bylo potvrdit skupiny podobných měst v rámci studie (Dobesova 2020). Cílem bylo nalezení podobností měst pomocí podobného výskytu a hodnoty podpory frekventovaných sad. Tyto podobnosti byly představeny v rámci kapitol 5.3 Případová studie – evropská města a 5.4 Porovnání s výsledky studie (Dobesova 2020).



Obrázek 10 Vývojový diagram postupu práce (zdroj: autor).

4 VLASTNÍ ŘEŠENÍ

Následující část práce se podrobně věnuje jednotlivým dílčím krokům zpracování práce. Klade si za cíl podrobně popsat, jak byla vybrána vstupní data, jak byla tato data zpracována a jakými postupy bylo dosaženo výsledků. Pro potřeby automatizace práce byly vytvořeny následující scripty a programy.

Program **SearchDistinctLanduse_SpatialJoin** pro výpočet transakčních dat sousednosti využití území, jehož součástí jsou dva scripty v jazyce Python **MakeDataDichotomous** a **ExportNonNullValues**. První zmíněný převádí data do dichotomického formátu (1 0), druhý exportuje všechny hodnoty nenabývající hodnoty Null (viz kapitola 4.3.3 SearchDistinctLanduse_SpatialJoin). Dále byly vytvořeny scripty **TXTtoEXCEL.py** a **Aggregation.py**, které modifikují vypočítané frekventované sady. TXTtoExcel.py vytváří z jednotlivých textových souborů přehledný soubor ve formátu MS Excel, který na každém listu obsahuje frekventované sady jednotlivých měst. Aggregation.py modifikuje sumarizační matici a odstraňuje z ní duplicity (viz kapitola 4.6 Příprava dat pro případovou studii – evropská města). Dále vznikl Jupyter Notebook **JoinTXT_Files.ipynb**, který spojuje transakční data několika měst do jednoho souboru. Tento Jupyter Notebook byl využit v rámci kapitoly (5.2 Případová studie – česká města).

4.1 Výběr dat

Práce se zaměřuje výhradně na data evropských měst, konkrétně na jejich městská jádra. Urban Atlas používá pojmy **Urban Core (městské jádro)** a **Functional Urban Area (funkční městská oblast)**, které se dále vyskytují v textu, proto je uvedena jejich definice.

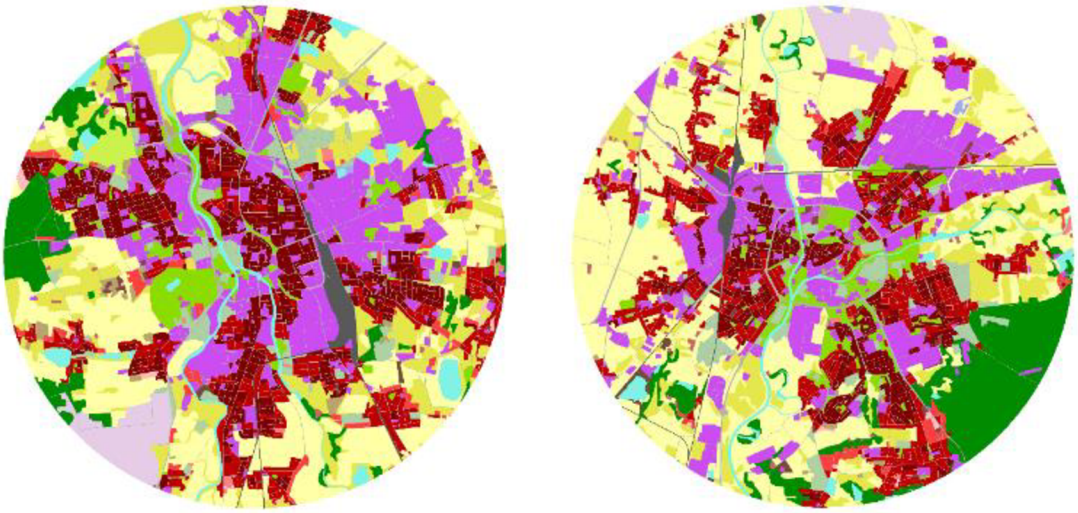
Městské jádro označuje hustě osídlenou centrální oblast města nebo metropolitní oblasti. Obvykle zahrnuje centrum města a okolní čtvrti s hustou zástavbou, obchodní čtvrti a kulturní či historické památky. **Funkční městská oblast (FUA)** je širší pojem, který zahrnuje městské jádro a okolní oblasti, které jsou s ním ekonomicky a sociálně integrovány. Zahrnuje dojížďkovou zónu městského jádra, kam lidé denně cestují za prací, studiem nebo za službami. FUA představuje spíše funkční jednotku pro plánování a tvorbu městské politiky než striktní administrativní hranici (Urban Atlas 2023).

V rámci datasetu Urban Atlas bylo vybráno **100 měst** pro další analýzu. Jako podklad pro výběr sloužil výběr (Janoušek 2019), který rovněž analyzoval 100 měst. Autor navrhl zaměření na města v rozmezí počtu obyvatel 50 000 – 200 000 obyvatel, protože se jedná o nejčetnější skupinu měst v rámci datasetu (ten čítá přibližně 800 měst). Rovněž bylo zahrnuto 22 měst ze studie Experiment in Finding Look-Alike European Cities Using Urban Atlas Data (Dobesova 2020). Jednalo se o dvojice měst (Obrázek 11), která byla identifikována jako podobná na základě zkoumání dat využití území datasetu Urban Atlas s využitím metody k-Nearest Neighbor nad feature vektorem získaným z neuronové sítě Painters (Kaggle 2016).

Celkový počet 100 se tedy skládá z 24 měst studie (Dobesova 2020) doplněných o 76 měst. Výběr měst byl proveden s ohledem na populační zastoupení států v Evropě. Zastoupení jednotlivých států bylo vypočítáno na základě počtu obyvatel a výsledné zastoupení menší než 1 bylo zaokrouhleno na 1, aby byla zajištěna přítomnost alespoň jednoho města z každého státu. Některé země, jako Německo, Francie, Itálie, Velká Británie, Polsko a Španělsko, byly ve finálním výběru zastoupeny méněkrát, než by odpovídalo jejich podílu počtu obyvatel, a to z důvodu nadhodnocení malých států a České republiky. Například Německo je zastoupeno 10 městy. Podrobně lze metodu přiřazení počtu států vidět v (Tabulka 3). Údaje o počtu obyvatel byly převzaty ze statistik Eurostat.

Tabulka 3 Přepočet podílu počtu obyvatel jednotlivých států v rámci datasetu na počet FUA při celkovém počtu 100.

Stát	Počet obyvatel	Podíl počtu obyvatel	Počet FUA (při celkovém počtu 100)	Počet FUA přepočteno	13 dvojic ze studie (Dobesova 2020)	Počet FUA CELKEM
Albania	2845955	0.0052	0.5157	1		**1**
Austria	9088681	0.0165	1.6468	2		**2**
Belgium	11730406	0.0213	2.1255	2		**2**
Bosnia and Herzegovina	3281000	0.0059	0.5945	1		**1**
Bulgaria	6948445	0.0126	1.2590	1	1	**1**
Croatia	4048165	0.0073	0.7335	1		**1**
Cyprus	1212754	0.0022	0.2197	1		**1**
Czechia	10703131	0.0194	1.9394	2	2	**8**
Denmark	5845165	0.0106	1.0591	1		**1**
Estonia	1328535	0.0024	0.2407	1		**1**
Finland	5546901	0.0101	1.0051	1		**1**
France	67413000	0.1221	12.2149	12	3	**8**
Germany	83927971	0.1521	15.2073	15	4	**10**
Greece	10741165	0.0195	1.9462	2		**2**
Hungary	9660351	0.0175	1.7504	2		**2**
Iceland	347897	0.0006	0.0630	1		**1**
Ireland	4952473	0.0090	0.8974	1		**1**
Italy	60367477	0.1094	10.9383	11	3	**9**
Kosovo	1806199	0.0033	0.3273	1		**1**
Latvia	1882809	0.0034	0.3412	1		**1**
Lithuania	2722289	0.0049	0.4933	1		**1**
Luxembourg	633700	0.0011	0.1148	1		**1**
Malta	514564	0.0009	0.0932	1		**1**
Montenegro	628066	0.0011	0.1138	1		**1**
Netherlands	17399993	0.0315	3.1528	3	3	**3**
North Macedonia	2077132	0.0038	0.3764	1		**1**
Norway	5448058	0.0099	0.9872	1		**1**
Poland	38369080	0.0695	6.9523	7	3	**6**
Portugal	10295000	0.0187	1.8654	2	1	**2**
Romania	19473936	0.0353	3.5286	3		**3**
Serbia	8737371	0.0158	1.5832	1		**1**
Slovakia	5462610	0.0099	0.9898	1		**4**
Slovenia	2100554	0.0038	0.3806	1	2	**2**
Spain	46954745	0.0851	8.5080	8	2	**6**
Sweden	10451295	0.0189	1.8937	2		**2**
Switzerland	8737795	0.0158	1.5832	2	2	**2**
United Kingdom	68207116	0.1236	12.3588	12		**8**
Celkem	***551891784***	***1.0000***	***100.0000***	***108***	***26***	***100***



(u) České Budějovice (Czech Republic) (v) Hradec Králové (Czech Republic)

Obrázek 11 Podobná města na základě analýzy ploch využití území a hierarchického shlukování převzato z: (Dobesova 2020).

Další zkoumanou skupinou měst byla česká města v rámci datasetu Urban Atlas 2018 (15 měst). V rámci těchto měst bylo cílem nalezení podobností na národní úrovni, ne mezi jednotlivými městy (5.2 Případová studie – česká města).

4.2 Zpracování vstupních dat

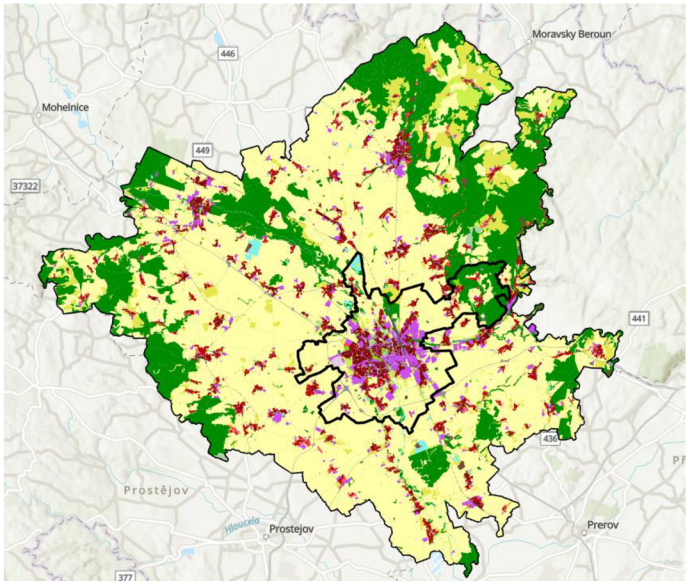
Na základě výběru byla stažena nejaktuálnější data Urban Atlas, a to konkrétně ta pro rok 2018 z webových stránek https://land.copernicus.eu/local/urban-atlas/urban-atlas-2018. Rozhraní bohužel neumožňuje stažení celé databáze najednou, uživatel musí zvolit, pro která města chce data stáhnout. Pro každé město obsahuje stažený archiv následující:

- **Data** – obsahuje geodatabázi ve formátu .gpkg
- **Documents** – obsahuje Delivery report a mapu
- **Legend** – obsahuje soubory legendy (.lyr, .qml)
- **Metadata** – obsahuje .xml soubor s metadaty

V rámci každé prostorové databáze se nachází tři třídy prvků v následujícím formátu pojmenování, který na příkladu města Praha je:

- CZ001L2_PRAHA_UA2018
- CZ001L2_PRAHA_UA2018_UrbanCore
- CZ001L2_PRAHA_UA2018_Boundary

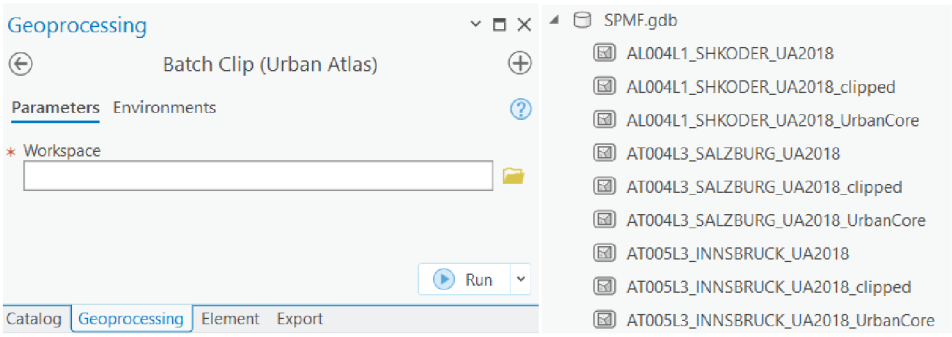
Třída prvků CZ001L2_PRAHA_UA2018 je polygonovou reprezentací využití území a zahrnuje celou oblast funkční urbánní oblasti (FUA – Functional Urban Area). Její hranici odpovídá třída prvků CZ001L2_PRAHA_UA2018_Boundary. Data v rámci celé datové sady dodržují stejné pojmenovávání jednotlivých vrstev, což významně zjednodušuje jejich zpracování. Pro potřeby práce jsou klíčové třídy prvků, jejichž název končí _UrbanCore. Jedná se o vymezení urbánních jader měst, se kterými je dále v práci počítáno.



Obrázek 12 Zdrojová data Urban Atlas 2018 pro Olomouc – vymezení Urban Core černým ohraničením (uprostřed).

Zdrojová data bohužel neobsahují třídu prvků, která by reprezentovala využití území pouze pro Urban Core. Pro potřeby práce byl vyvinut **script v jazyce Python – Batch Clip (Urban Atlas) v prostředí ArcGIS Pro** (příloha: Vystupni_Data/PROJECT_ARCGISPRO/MAIN/MAIN.atbx), který slouží k hromadnému ořezání vstupních dat pouze na oblast Urban Core. Vzniklý nástroj pracuje s konzistentním pojmenováním v rámci celého datasetu. Vstupem je Workspace (geodatabáze), ve které se nachází všechna vstupní data (100 polygonů FUA, 100 polygonů hranice FUA a 100 polygonů hranice Urban Core). Script si vytváří seznam všech tříd prvků a hledá prvek, jehož název končí řetězcem "_UrbanCore". K němu hledá korespondující prvek. Nástroj dále provede oříznutí dvojice dat a výsledek pojmenuje s příponou "_clipped". Výsledkem je tedy 100 nových tříd prvků, které vstupují do dalšího zpracování. Nástroj přináší do procesu zpracování prvek automatizace. Klasický nástroj Clip disponuje Batch verzí, ta však umožňuje pouze pro několik vstupních tříd prvků nastavit jednu konkrétní překryvnou vrstvu. Z tohoto důvodu byl implementován jednoduchý script (Obrázek 13).

Výstupem této fáze zpracování jsou dvě geodatabáze v ArcGIS, a to konkrétně databáze **Europe_UA2018** – obsahující 100 měst a databáze **Czechia_UA2018** – obsahující 15 českých měst. Tato data dále vstupují do následující fáze zpracování.



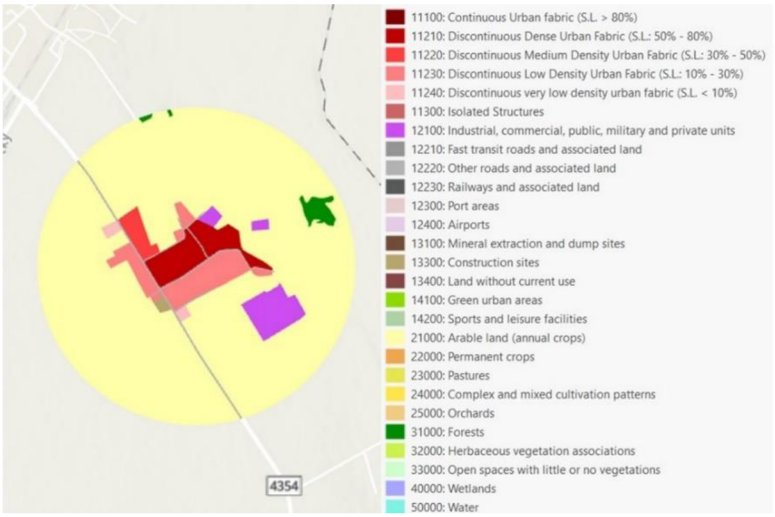
Obrázek 13 Nástroj pro hromadné ořezání vstupních dat a náhled na databázi se zpracovanými vstupními daty.

```
Import arcpy
# get workspace parameter from user input
workspace = arcpy.GetParameterAsText(0)
# set workspace to user-specified value
arcpy.env.workspace = workspace
# list all 34eatures classes in workspace
fcs = arcpy.ListFeatureClasses()
# loop through each 34eatures class in workspace
for fc in fcs:
    # check if 34eatures class ends with „_UrbanCore“
    if fc.endswith(„_UrbanCore“):
        # construct name of corresponding 34eatures class
        fc_name = fc[:-10] # remove „_UrbanCore“
        # check if corresponding 34eatures class exists
        if arcpy.Exists(fc_name):
            # set input and clip features
            in_features = fc_name
            clip_features = fc
            # set output 34eatures class name
            out_feature_class = fc_name + „_clipped“
            # perform clip analysis
             arcpy.Clip_analysis(in_features,clip_features,
             out_feature_class)
```

4.3 Implementace nástroje pro přípravu kategoriálních a dichotomických dat

Stěžejní částí práce je implementace nástroje pro přípravu kategoriálních a dichotomických dat. Žádný z programů uvedených v kapitole 3.1 Použité programy neumožňuje přímo pracovat s prostorovými daty. Data miningové programy vyžadují specificky formátovaná data, v našem případě ve formě transakčních tabulek a textových souborů. Koncept transakčních dat je blíže představen v kapitole 2.3 Frekventované sady. Výstupem nástroje je transakční tabulka všech jedinečných kódů využití území nacházejících se do uživatelem nastavené vzdálenosti.

Funkcionalita je blíže představena na vesnici Štětovice (Obrázek 14), část obce Vrbátky v okrese Prostějov. Jedná se o malou vesnici, jež je součástí datasetu FUA Olomouc. Nachází se asi 1,5 km na jihovýchod od Vrbátek. V roce 2009 zde bylo evidováno 132 adres. Vesnice byla vyříznuta pomocí kruhové výseče a je reprezentována 18 polygony. Jedná se o velmi malou reprezentaci reálných vstupních dat, na které je možno ručně kontrolovat výstupy jednotlivých nástrojů. Rovněž se jedná o ideální reprezentaci pro představení funkcionality a architektury jednotlivých nástrojů.



Obrázek 14 Vesnice Štětovice v rámci FUA Olomouc, trénovací data pro ladění nástrojů.

V rámci práce byly připraveny 3 varianty nástroje, které se liší jak samotnou logikou postupu zpracování, tak výpočetní rychlostí. V následující části jsou postupně představeny.

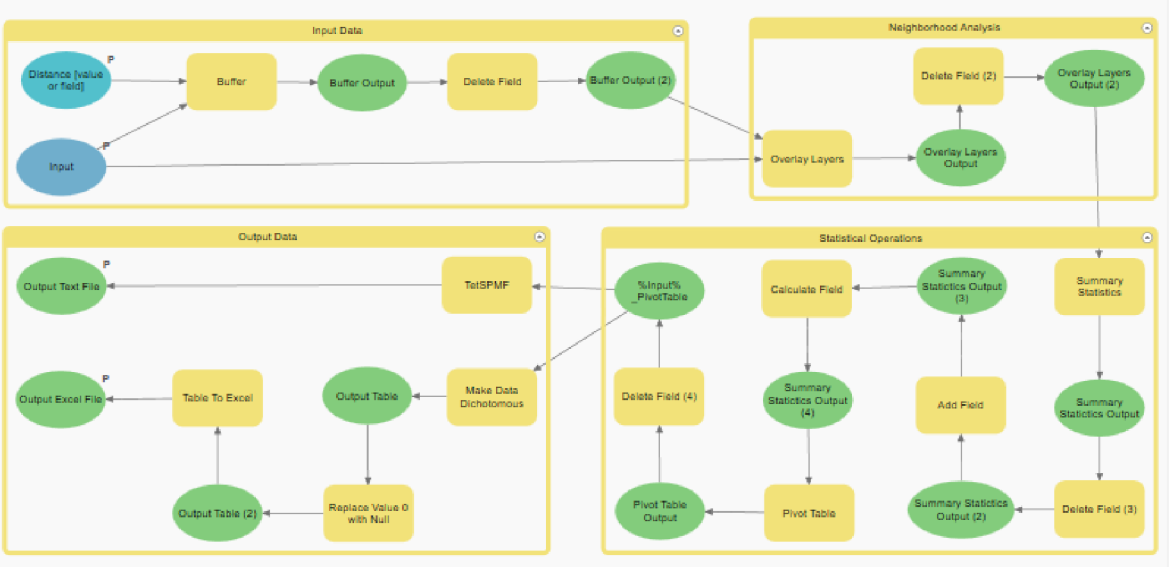
4.3.1 SearchDistinctLanduse_GenerateNearTable

Jedná se o prvotní návrh na řešení problematiky. Nástroj je vytvořen v prostředí ModelBuilder v ArcGIS Pro. Používá kombinaci nástrojů Spatial Join, **Generate Near Table** a Summary Statistics. Jeho hlavní nevýhodou a také důvodem, proč nebyl využit pro zpracování dat, byla jeho neschopnost přizpůsobení se tvaru vstupních polygonů. To bylo způsobeno funkcionalitou nástroje Generate Near Table, která hledá sousední polygony v zadané vzdálenosti.

U podlouhlých vstupních polygonů dochází ke generování sousedností ve větším počtu v jednom směru než v ostatních směrech. Nástroj vypíše do tabulky všechny sousedící polygony podle centroidů, což je další logický problém, protože podlouhlý polygon může mít centroid ve větší vzdálenosti, než je prahová vzdálenost nástroje.

4.3.2 SearchDistinctLanduse_OverlayLayers

Druhou iterací byl nástroj využívající funkcionalitu nástroje **Overlay Layers** (Obrázek 15) v rámci ArcGIS Pro. Nástroj nejprve vytvoří buffer s uživatelem nastavenou vzdáleností. Tento buffer společně se vstupní vrstvou vstupuje do nástroje Overlay Layers. Ten překryje jednotlivé geometrie a zkombinuje jejich atributy. Následně byl použit nástroj Summary Statistics. Ten vypočítal počet jednotlivých typů landuse v okolí. Celkově se nástroj ukázal jako nepoužitelný pro velké vstupní datasety, protože tvoří velké dočasné soubory v řádu desítek GB u měst velikosti 200 000 obyvatel.



Obrázek 15 Nástroj SearchDistinctLanduse_OverlayLayers.

4.3.3 SearchDistinctLanduse_SpatialJoin

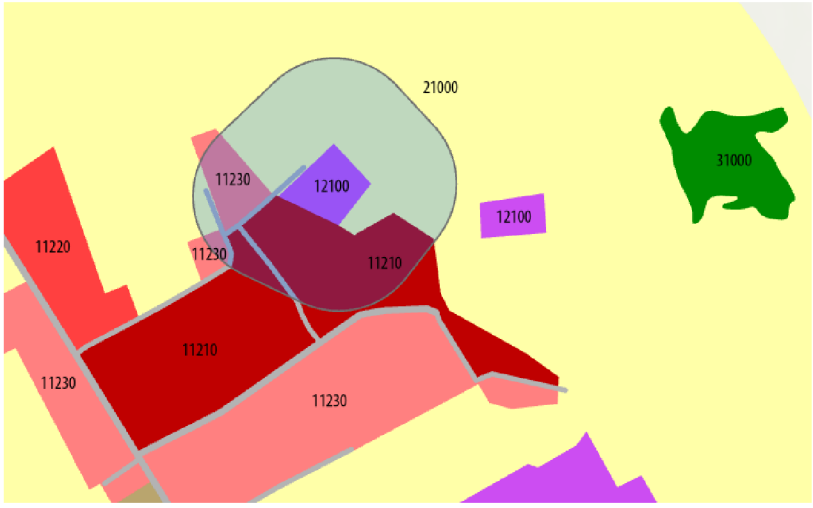
V rámci třetí iterace byl vytvořen nástroj pracující na principu Spatial Join. Model lze vidět na (Obrázek 21). Jedná se o kombinaci geoprocesingových nástrojů a Python skriptů. Jeho funkčnost je následující:

1. Vstupními daty je polygonová data z Urban Atlas 2018.
2. Prvním nástrojem v modelu je **Buffer**, jehož vzdálenost je nastavitelná uživatelem.
3. Následně jsou z výsledného bufferu odstraněny pomocí nástroje **Delete Field** všechna pole, aby nedocházelo ke konfliktu při následných operacích (jak zdrojová data, tak buffer obsahuje atribut "code_2018", což je nežádoucí).
4. Dále je použit nástroj **Spatial Join** s nastavením:
    - Target Feature = výstup nástroje Buffer
    - Join Feature = vstupní polygonová data Urban Atlas 2018
    - Join Operation = Join one to many
    - Match Option = Intersect

Výsledkem této operace je vrstva, která pro každý vstupní polygon (Join Feature), který se protne s bufferem (Target Feature), vytvoří jeden záznam (řádek) s údajem o využití území v originální vrstvě. V tomto případě pro buffer polygonu TARGET_FID=6 s landuse 12100 vytvoří následující záznamy (Obrázek 16). Výsledek nám říká, že polygon č. 6 vstupní vrstvy má do vzdálenosti 100 m následujících 7 využití území jako své sousedy (11210, 11210, 12100, 11230, 11230, 12220, 21000). Správnost výstupu můžeme vidět na (Obrázek 17). V tuto chvíli výsledky obsahují duplicity a je potřeba je odstranit.

	fid *	geom *	Join_Count	JOIN_FID	TARGET_FID	code_2018	country	fua_name	fua_code
34	34	Polygon Z	1	1	6	11210	CZ	Olomouc	CZ006L2
35	35	Polygon Z	1	5	6	11210	CZ	Olomouc	CZ006L2
36	36	Polygon Z	1	6	6	12100	CZ	Olomouc	CZ006L2
37	37	Polygon Z	1	7	6	11230	CZ	Olomouc	CZ006L2
38	38	Polygon Z	1	8	6	11230	CZ	Olomouc	CZ006L2
39	39	Polygon Z	1	12	6	12220	CZ	Olomouc	CZ006L2
40	40	Polygon Z	1	17	6	21000	CZ	Olomouc	CZ006L2

Obrázek 16 Výstup nástroje Spatial Join.



Obrázek 17 Vizuální kontrola správnosti generování sousedních využití půdy.

5. Data jsou dále zpracována nástrojem **Pivot Table** s nastavením:
   - Input Field = TARGER_FID
   - Pivot Field = "code_2018"
   - Pivot Value = "code_2018"

Pomocí nastavení je docíleno, že pro každé TARGET_FID, což reprezentuje jeden vstupní buffer je vytvořen jeden řádek v nové tabulce. Z existujících "code_2018", které reprezentují existující využití území v datasetu, jsou vytvořeny nové sloupce, do kterých je vložena hodnota jednotlivých landuse. Pomocí nástroje jsou odstraněny duplicity. Každá hodnota landuse, která se v bufferu nachází je zaznamenána právě jednou (Obrázek 18).

	fid *	TARGET_FID	11210	11220	11230	11240	12100	12220	13300	21000	31000
1	1	1	11210	11220	11230	<Null>	12100	12220	13300	21000	<Null>
2	2	2	11210	<Null>	11230	<Null>	12100	12220	<Null>	21000	<Null>
3	3	3	11210	<Null>	<Null>	<Null>	12100	12220	<Null>	21000	<Null>
4	4	4	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	<Null>	21000	31000
5	5	5	11210	<Null>	11230	<Null>	12100	12220	<Null>	21000	<Null>
6	6	6	11210	<Null>	11230	<Null>	12100	12220	<Null>	21000	<Null>

Obrázek 18 Výstup nástroje Pivot Table.

6. V závěrečné části nástroje dochází ke **generování dvou typů výstupů**. Jedním z nich je **textový soubor** s názvem %Input%_FreqItemsetData.txt jež je generován scriptem v jazyce Python (ExportNonNullValues.py) vytvořeným pro tuto práci. Script zpracovává výstup nástroje Pivot Table (Obrázek 18). Script filtruje pryč všechny hodnoty NULL a také odstraňuje pole s ID hodnotou a do výsledného textového souboru zapisuje pouze jednotlivé hodnoty využití landuse oddělené mezerami (Obrázek 19). **Výstupní soubor tak obsahuje na každém řádku data jedné transakce.** Modře zvýrazněn je polygon č. 6, který byl v předchozích krocích použit k vysvětlení.

```
import arcpy
# Define parameters
input_table = arcpy.GetParameterAsText(0)
output_file = arcpy.GetParameterAsText(1)
# Open output file for writing
with open(output_file, ,w') as f:
    # Loop through rows in input table
    with arcpy.da.SearchCursor(input_table, ,*') as cursor:
        for row in cursor:
            # Filter out null values and ID field
            row_values = [str(val) for val in row[1:] if val is not None]
            # Write row values to output file
            f.write(, ,.join(row_values) + ,\n')
```

Stetovice_FreqItemsetData.txt - Notepad

File Edit Format View Help

```
11210 11220 11230 12100 12220 21000
11210 11230 12100 12220 21000
11210 12100 12220 21000
21000 31000
11210 11230 12100 12220 21000
11210 11230 12100 12220 21000
11210 11220 11230 12100 12220 21000
11210 11230 12100 12220 21000
11210 11220 11230 11240 12220 21000
11210 11220 11230 11240 12220 21000
11220 11230 11240 12220 21000
11210 11220 11230 11240 12100 12220 21000
11210 11230 11240 12100 12220 21000
11230 11240 12220 21000
11210 11230 11240 12220 21000
11210 11220 11230 11240 12100 12220 21000 31000
21000 31000
11210 11220 11230 11240 12220 21000
```

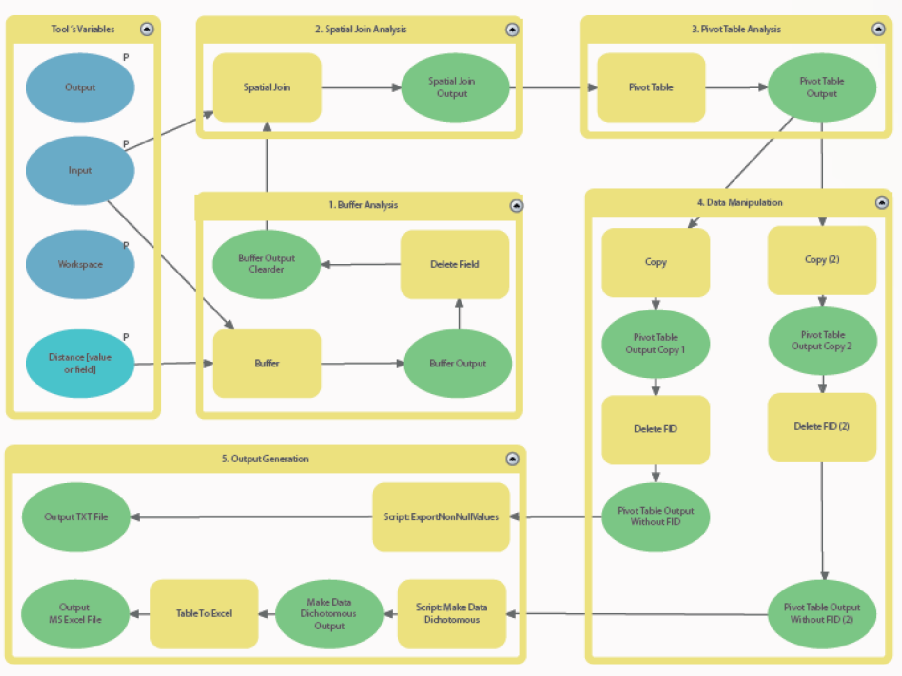
Obrázek 19 Výstup nástroje SearchDistinctLanduse_SpatialJoin – textový soubor.

Druhým výstupem je **soubor MS Excel** (Obrázek 20) obsahující dichotomická data (pro použití v programu Orange (hodnoty 0 a 1). Pro vygenerování tohoto výstupu byl implementován další script v jazyce Python **Make Data Dichotomous** příloha: Vystupni_Data/PROJECT_ARCGISPRO/MAIN/SearchDistinctLanduse_SpatialJoin.atbx).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
OBJECTID	11100	11210	11220	11230	11240	11300	12100	12220	13100	13400	14100	14200	21000	22000	23000
1	0	1	1	1	0	0	1	1	0	0	0	0	1	0	0
2	0	1	0	1	0	0	1	1	0	0	0	0	1	0	0
3	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
5	0	1	0	1	0	0	1	1	0	0	0	0	1	0	0
6	0	1	0	1	0	0	1	1	0	0	0	0	1	0	0

Obrázek 20 Výstup nástroje SearchDistinctLanduse_SpatialJoin – soubor MS Excel soubor.

```
import arcpy
# Get input table from modelbuilder
input_table = arcpy.GetParameterAsText(0)
# Get output table name and location from modelbuilder
output_table = arcpy.GetParameterAsText(1)
# Create feature layer from input table
feature_layer = arcpy.MakeTableView_management(input_table, "feature_layer")
# Create list of field names (excluding ORIG_FID) and update cursor
field_list = arcpy.ListFields(input_table)
field_names = [field.name for field in field_list if field.name != "ORIG_FID"]
with arcpy.da.UpdateCursor(feature_layer, field_names) as cursor:
    for row in cursor:
        # Replace null values with 0, non-null values with 1
        row = [0 if value is None else 1 for value in row]
        cursor.updateRow(row)
# Use the CopyRows tool to copy the updated data to the new table
arcpy.CopyRows_management(feature_layer, output_table)
# Delete the feature layer
arcpy.Delete_management(feature_layer)
```



Obrázek 21 Nástroj SearchDistinctLanduse_SpatialJoin.

Z výše zmíněných 3 nástrojů lze první dva považovat za testovací a třetí za reálně implementovaný. Z výše zmíněných důvodů byly první dva nástroje opuštěny. Třetí nástroj byl nakonec převeden do Batch formy umožňující zpracovávat více vstupních dat najednou. Nástroj byl následně aplikován na specifikovaná data z kapitoly (4.1 Výběr dat).

- Data pro případovou studii evropských měst (100)
- Data pro případovou studii českých měst (15)

Výstupem této části práce jsou transakční data sousednosti jednotlivých polygonů landuse pro jednotlivá města, a to jak ve formě textového kategoriálního souboru, tak souboru MS Excel s dichotomickými daty.

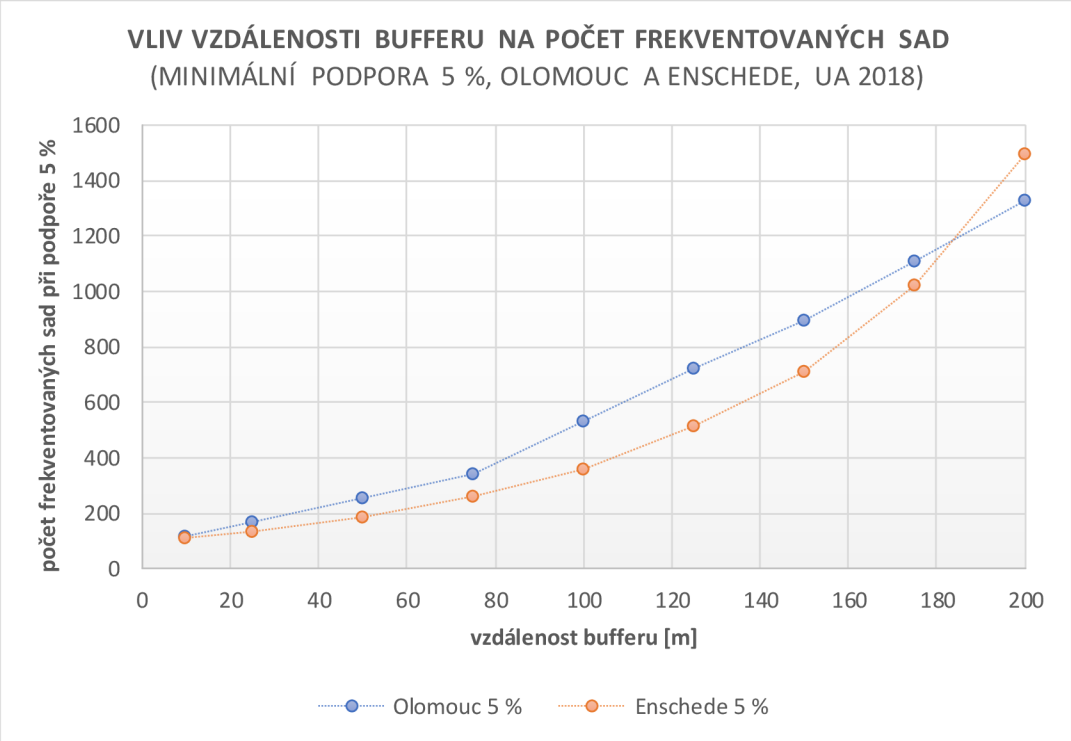
4.4 Nastavení vzdálenosti nástroje

Hlavním parametrem nástroje SearchDistinctLanduse_SpatialJoin je vzdálenost bufferu. Vzdálenost ovlivňuje množství polygonů vstupujících do zjišťování sousednosti. Obecně platí, že čím je vzdálenost větší, tím více má polygon sousedů a tím delší jsou jednotlivé transakce. Při nastavování vzdálenosti bufferu je potřeba brát v potaz **minimální mapovatelnou jednotku** datové sady Urban Atlas (Tabulka 4). Ta je u **tříd 1 (urbánní třídy): 0,25 ha** a u **tříd 2–5 (rurální třídy): 1 ha**. To je rovno straně 50 respektive 100 metrů. Nastavená vzdálenost by tedy neměla být menší než 100 m.

Tabulka 4 Přesnost produktu Urban Atlas 2018. (zdroj: land.copernicus.eu/user-corner/technical-library/urban-atlas-mapping-guide)

	Třídy CORINE	**Úrovně třídy**	**Minimální mapovací jednotka**	**Tematická přesnost**	**Prostorová přesnost pixelu**
Urbánní	1	I–IV	0,25 ha	≥ 85%	+/- 5 m
Rurální	2–5	I–II	1,00 ha	≥ 80%	+/- 5 m
Celková přesnost				≥ 80%	

V této fázi zpracování dat byly experimentálně zjišťovány frekventované sady pro různou vzdálenost bufferu kolem vstupních polygonů landuse. Byly vygenerovány sousednosti pro rozsah 10 až 200 m pro data města Olomouce a Enschede. Počet frekventovaných sad při podpoře 5 % v závislosti na velikosti bufferu ukazuje (Obrázek 22).



Obrázek 22 Vliv vzdálenosti bufferu na počet frekventovaných sad (minimální podpora 5 %, Olomouc, UA 2018).

Bylo zjištěno, že nastavená vzdálenost bufferu, má vliv také na **mediánovou délku transakce**. Jedná se o medián (střední hodnotu) počtu prvků v jednotlivých frekventovaných sadách. Medián byl použit z důvodu minimalizování vlivu extrémně dlouhý transakcí, které vznikají z důvodu struktury vstupních dat Urban Atlas. Např. polygony třídy silnice jsou v rámci datasetu reprezentovány velkými polygony, které se rozprostírají přes celé zájmové území.

U takového vstupního polygonu při aplikování nástroje dochází k nalezení téměř všech kategorií v rámci datasetu, a tudíž i neúměrně dlouhé transakci. Takováto transakce nemá na nalezení výsledků významný vliv, protože takto dlouhá transakce je vyhodnocena jako nefrekventovaná, a tudíž nevstupuje do vyhodnocení. Naopak v případě výpočtu sousednosti z jiných malých polygonů netvoří velké polygony silnic problém, protože jsou překryty bufferem malého polygonu. Obecně, **čím je vzdálenost větší, tím je mediánová délka transakce větší** (Tabulka 5).

Tabulka 5 Vliv vzdálenosti bufferu na průměrnou a mediánovou délku transakce – městské jádro Olomouc (UA 2018).

Vzdálenost bufferu	**Průměrná délka transakce**	**Mediánová délka transakce**
10	4.26923077	4
25	4.55649038	4
50	4.92908654	5
75	5.29206731	5
100	5.65444712	5
125	5.98918269	6
150	6.28906250	6
175	6.55649038	6
200	6.83653846	7

Ve fázi experimentování s velikostí bufferu byla nastavena i vzdálenost na 400 metrů a napočítána transakční data sousednosti pro město Olomouc. Tento návrh byl inspirován studiemi, které se zaměřily na prahové vzdálenosti pro docházkovou vzdálenost k zastávkám hromadné dopravy (Burian et al. 2016) a vzdálenost od bydliště. Pokud bychom takový návrh aplikovali, bylo by pravděpodobné, že transakce by obsahovaly vlastní bydliště, místo pro volnočasové aktivity (nejbližší zeleň), obchody a další služby, které jsou v docházkové vzdálenosti. Nicméně, vzhledem k metodě frekventovaných sad se ukázalo, že tento návrh nepřináší podstatná zjištění. S rostoucí vzdáleností bufferu skokově narůstá počet unikátních frekventovaných sad v rámci města, které ale mají velmi malou podporu napříč městem a v konečném důsledku i mezi jednotlivými městy. Respektive se tyto frekventované sady nedostanou nad hranici minimální podpory, a tudíž se neberou v úvahu při vyhodnocování. Ve výsledku převáží frekventované sady sousedností, které jsou ve větší blízkosti. Tato situace také brání efektivnímu nalezení podobností, což je jeden z cílů této práce.

Po vyhodnocení experimentů byla vzdálenost nastavena na finální hodnotu 100 m. S touto hodnotou je potřeba pracovat při interpretaci výsledků. Intepretace dat sousednosti je následující: **jedna transakce v rámci transakčních dat reprezentuje všechna unikátní (neopakující se) využití území v okolí vstupního polygonu do vzdálenosti 100 m včetně vstupního polygonu.** Jakékoliv analýzy podobnosti v následující části

práce pracuje se vzdáleností 100 m, a tudíž je potřeba tuto vzdálenost použít při interpretování výsledků. Jakákoliv podobnost nalezená na základě sousednosti se tedy vztahuje na tuto vzdálenost.

4.5 Výpočet frekventovaných sad

Na základě dat získaných v rámci kapitoly 4.3.3 SearchDistinctLanduse_SpatialJoin bylo provedeno generování frekventovaných sad. Vstupní data byla v následujících formátech: textový soubor a soubor MS Excel. Pro generování byly zvažovány 3 programy a to Orange, SPMF a implementace algoritmu negFIN v Jupyter Notebooku (FI.ipynb). Z důvodů popsaných v kapitole (3.1 Použité programy) nebylo dále pracováno v programu Orange. Pro shrnutí, program Orange neumožňuje export výsledků do strojově čitelné podoby a nelze s nimi dále efektivně pracovat (fitrovat, řadit atd.).

Vstupní textový soubor z nástroje pro generování sousednosti je ve formátu, který je zpracovatelný programem SPMF a Jupyter Notebook. Program **SPMF** byl použit pro testování jednotlivých algoritmů a nastavení. Primárně byl použit **mód Pattern Viewer**, který umožňuje interaktivně prohlížet výsledky, filtrovat je a dále exportovat pouze výběry dat. Hlavním nedostatkem programu je výpočet podpory frekventovaných sad v absolutních hodnotách (#SUP). Absolutní hodnoty podpory vyjadřují četnost transakcí ale zároveň i závisí na celkovém počtu transakcí ve zdrojových datech. Dvě sady transakcí (zde dvě města) s různým počet transakcí a s různou skladbou transakcí nelze mezi sebou porovnávat na základě absolutní hodnoty podpory. Výsledky z programu SPMF obsahují pouze absolutní hodnoty podpory a bylo by nutné je mimo program SPMF přepočítat pro následné porovnání dvou sad transakcí (v řešeném případě dvou měst).

Hlavním nedostatkem programu je **výpis pouze absolutní podpory jednotlivých frekventovaných sad**. Před používáním implementace v Jupyter Notebooku bylo uvažováno dopočítávání relativní podpory v rámci postprocesingu dat. Zde by docházelo k jednoduchému přepočtu na relativní četnost. Od toho bylo upuštěno a SPMF dále figuroval pouze jako program pro interaktivní výpočty a generování mezivýstupů, které do další práce nevstupovaly.

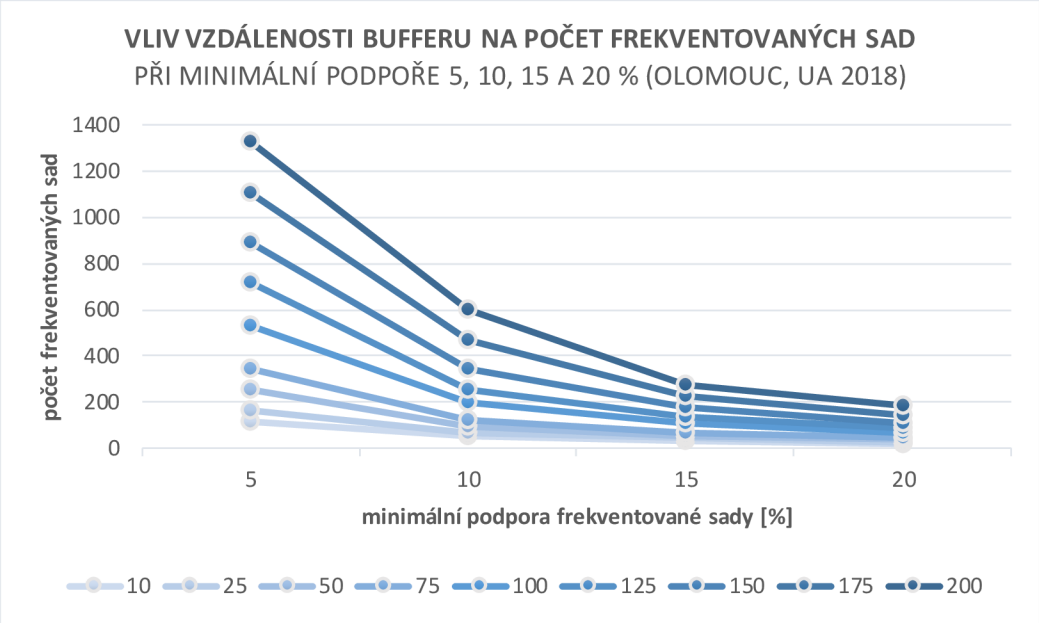
Po vyřazení programu Orange bylo přistoupeno ke pracování v programu **FI.ipynb** vyvinutého v rámci spolupráce s Univerzitou v Kyoto. Jedná se o program v prostředí Jupyter Notebook umožňující hromadné zpracování dat. V následujících podkapitolách je představeno využití jednotlivých programů a postup zpracování dat.

4.5.1 Závislost počtu frekventovaných sad na minimální podpoře

Vztah mezi počtem frekventovaných sad a minimální podporou je důležitým konceptem v dolování dat. Minimální podpora se nastavuje až při samotném generování frekventovaných sad, jak je popsáno v kapitole 4.5 Výpočet frekventovaných sad. Obecně platí, že s rostoucím prahem minimální podpory klesá počet frekventovaných sad (sad, které splňují nebo překračují práh minimální podpory). A naopak, s klesajícím prahem minimální podpory se počet frekventovaných sad zvyšuje.

Například pokud máme soubor dat se 100 transakcemi a nastavíme práh minimální podpory na 10 %, což znamená, že se množina položek musí objevit alespoň v 10 transakcích, aby byla považována za častou, můžeme nakonec získat jen několik frekventovaných sad. Pokud však snížíme minimální práh podpory na 5 %, můžeme najít mnohem více sad, včetně těch, které nebyly identifikovány při použití vyššího prahu.

Na (Obrázek 23) je vidět, že při minimální podpoře 20 % se vliv vzdálenosti bufferu výrazně potlačuje, počet pravidel je mezi do 200. Zatímco při podpoře 5 % je počet frekventovaných sad výrazně vyšší pro větší buffer.



Obrázek 23 Vliv vzdálenosti bufferu na počet frekventovaných sad při minimální podpoře 5, 10, 15 a 20 % (Olomouc, UA 2018).

Obecně platí, že čím vyšší je minimální práh podpory, tím selektivnější je proces dolování, což vede k menšímu počtu častých množin položek. **Příliš vysoké nastavení prahu minimální podpory** však může způsobit, že nám **uniknou některé potenciálně zajímavé nebo užitečné vzory**. Na druhou stranu **příliš nízké nastavení prahu minimální podpory** může vést k **nadměrnému počtu frekventovaných sad**, což ztěžuje získání smysluplných poznatků. Volbu minimální prahové hodnoty podpory je proto třeba pečlivě zvážit na základě konkrétního souboru dat a daných výzkumných otázek.

4.5.2 Jupyter Notebook

V rámci spolupráce s Univerzitou Kyoto a panem Tai Dinh nám byla poskytnuta implementace algoritmu negFIN (Aryabarzan et al. 2018), který představuje odklonění se od původní myšlenky použití algoritmu Apriory (Agrawal a Srikant 1994). Algoritmus negFIN je rovněž implementován v programu SPMF.

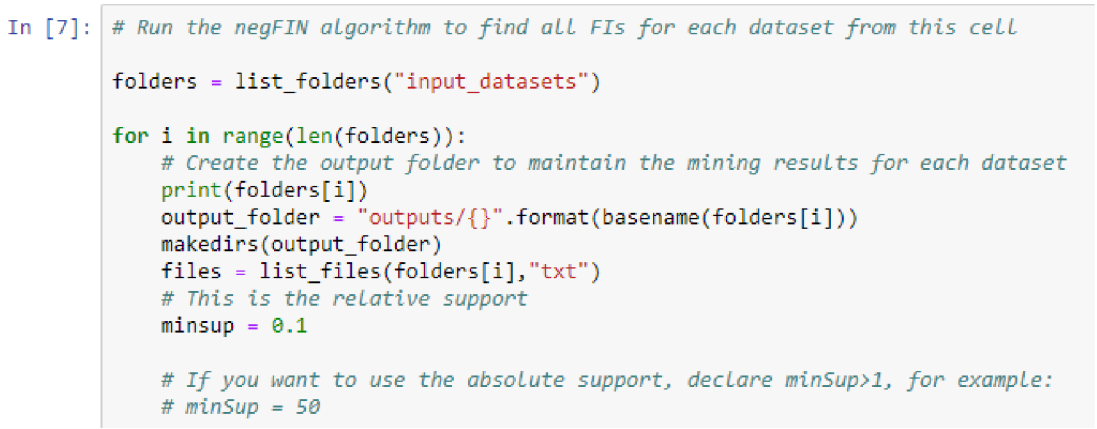
Jupyter Notebook (Obrázek 24) je webový interaktivní editor, který může obsahovat jednak scripty v jazyce Python, dále potom strukturovaný text, matematické vzorce (LaTeX) a další prvky. Rozhraní se zobrazuje ve webovém prohlížeči. Notebook pracuje v rámci složky, ze které je spuštěn. Každý Jupyter Notebook se skládá z jednotlivých buněk, které mohou obsahovat jednotlivé části kódu.

Na základě konzultace s vedoucí práce byla použita **minimální podpora frekventovaných sad 5 %**. Tato hodnota generuje velké množství frekventovaných sad a zároveň odfiltrovává velké množství velmi málo frekventovaných sad. Vliv nastavení minimální podpory byl diskutován v rámci kapitoly 4.5.1 Závislost počtu frekventovaných sad na minimální podpoře.

Obrázek 24 Náhled webového rozhraní Jupyter Notebook – ukázkový kód.

V rámci pracovního adresáře poskytnutého programu byly vytvořeny složky input_datasets a outputs. Do složky input_datasets byly vloženy napočítané textové soubory z přechozího zpracování (nástroj: SearchDistinctLanduse_SpatialJoin). Výstupem jsou soubory s napočítanými frekventovanými sadami. Tento přístup je významným krokem k automatizaci práce. Na jednu stranu nelze výsledky interaktivně prohlížet (výstupy jsou textové soubory) ani filtrovat bez dalšího softwaru. Na druhou stranu lze tyto textové soubory dále pomocí následných scriptů zpracovávat.

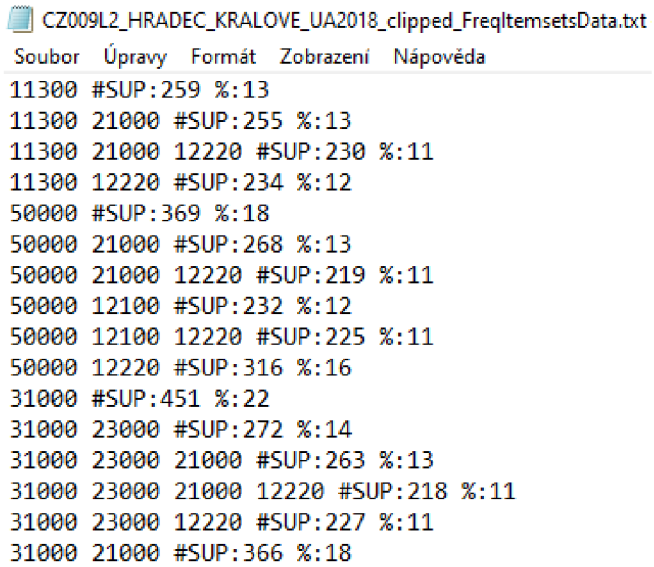
Poskytnutý Jupyter Notebook je součástí příloh. Minimální podporu vypočítaných pravidel lze specifikovat v buňce č. 7 v rámci proměnné **minsup** (Obrázek 25). Napočítané frekventované sady pro jednotlivá města ve formě textových souborů lze nalézt ve složce outputs. Takto zpracovaná města vstupují do další fáze zpracování.



Obrázek 25 Buňka Jupyter Notebooku FI.ipynb s možností nastavení minimální podpory pravidel (zdroj: Tai Dinh, University of Kyoto).

4.6 Příprava dat pro případovou studii – evropská města

V rámci první případové studie byla zkoumána vybraná evropská města. Výběr byl blíže představen v kapitole 4.1 Výběr dat. Pro všechna města hromadně byly vypočítány frekventované sady. Výstupem výpočtu bylo 100 textových souborů. Na příkladu Hradce Králové (Obrázek 26) můžeme vidět výčet prvních 16 frekventovaných sad z celkového počtu 147. Název souboru je kombinací názvu vstupních dat městského jádra UA2018 "CZ009L2_HRADEC_KRALOVE_UA2018_clipped" a přípony "_FrequentItemsetsData.txt". Textový soubor obsahuje všechny frekventované sady, které splňují kritérium minimální podpory 5 %.



Obrázek 26 Textový soubor s vypočítanými frekventovanými sadami – Hradec Králové, minimální podpora 5 %, výběr hodnot.

Tato data byla dále zpracována do podoby matice, jež má následující strukturu:

- **Sloupce** – výčet všech unikátních (neopakujících se) frekventovaných sad
- **Řádky** – výčet všech vstupních měst
- **Jednotlivé hodnoty** – procentuální podpora jednotlivých frekventovaných sad v daném městě

Matice byla vytvořena pomocí scriptu v jazyce Python. Script je součástí Jupyter Notebooku. Skript provádí následující úlohy:

- V prvním kole prochází města a získává seznam názvů měst a FI (frequent itemsets). Soubor se prohledá a každý FI se vloží do seznamu FI_list a poté se vypíše.
- Poté se odstraní duplicity FI v seznamu FI_list a všechny FI v seznamu unique_FI_list se seřadí podle abecedy a podle délky.
- Posledním krokem je vytvoření datového rámce s city_list pro řádky a unique_FI_list pro sloupce.
- Ve druhém kole se opět prochází města, tentokrát za účelem doplnění hodnot (procentuální podpory jednotlivých pravidel). Soubor se prohledá a hodnoty se doplní do datového rámce.
- Nakonec se datový rámec uloží do výstupního souboru MS Excel.

```
def summarize_result(input_folder,cities):
    # FIRST ROUND: traverse cities to get the list of city names (they will be rows
      in the data frame) and FIs (they will be columns in the data frame)
    FI_list = []
    city_list = []
    city_index = []
    files = list_files(input_folder,"txt")
    for i in range(len(files)):
        file_name = basename(files[i])
        start_pos = file_name.find("_") + 1
        end_pos = file_name.find("_UA2018")
        city_name = file_name[start_pos:end_pos].rsplit('_', 2)[-3:]
        city_name = '_'.join(city_name)
        if cities == 'all_city' or city_name in cities:
            city_list.append(city_name)
            city_index.append(i)
            # scan the file
            with open(files[i], 'r') as reader:
                for line in reader:
                    # Put the FI into the FI_list
                    sup_pos = line.find("#SUP")
                    FI = line[:sup_pos].strip()
                    # print(FI)
                    FI_list.append(FI)
    print("Total of cities:",len(city_list))
    print("Total of FIs in {} cities:{}".format(len(city_list),len(FI_list)))
    # Remove the duplication of FI in the FI_list
    unique_FI_list = list(set(FI_list))
    print("Total    of    FIs    in    {}    cities    after    remove    the
duplicate:{}".format(len(city_list),len(unique_FI_list)))
    # Sort all FIs in the unique_FI_list according to their name and length
    unique_sorted_FI_list = sorted(unique_FI_list, key=lambda s: (len(s), s))
    # Create a data frame that get city_list for rows and unique_FI_list for columns
    df_city_FI = pd.DataFrame(0, index=city_list, columns=unique_sorted_FI_list)
    print(df_city_FI.shape)
    # SECOND ROUND: traverse cities to fill in values
    city_index_name_dict = dict(zip(city_index,city_list))
    for i in city_index:
        # scan the file
        with open(files[i], 'r') as reader:
            for line in reader:
                sup_pos = line.find("#SUP")
                FI = line[:sup_pos].strip()
                percentage = line.split('%:')[1].strip()
                # Fill the values
                df_city_FI.loc[city_index_name_dict[i], FI] = percentage
    # Save the dataframe into an excel output file
    if cities == 'all_city':
        df_city_FI.to_excel('summary_all_cities.xlsx')
```

```
    else:
        file_name = '-'.join(cities) + '.xlsx'
        df_city_FI.to_excel(file_name)
    print("Done!")
# Summary of all cities in folder outputs \Europe_UA2018_100m
summarize_result("outputs\Europe_UA2018_100m", 'all_city')
# Summary of the result for three cities SHKODER and SALZBURG and USTI_NAD_LABEM
summarize_result("outputs\Europe_UA2018_100m",
['SHKODER','SALZBURG','USTI_NAD_LABEM'])
# Summary of the result for only one selected city (e.g. SHKODER in Albania)
   summarize_result("outputs\Europe_UA2018_100m", ['SHKODER'])
```

Vzniklá matice ale bohužel obsahovala duplicity. Program např. nepovažoval frekventované sady „11100 12000“ a „12000 11100“ za totožné. Proto byl v rámci práce implementován další script **Aggregation.py** (přílohy: Vystupni_Data/SCRIPTY/Aggregation/Aggregation.py), jež vzniklé duplicity odstranil. Skript načte soubor aplikace Excel s názvem 'input.xlsx' do datového rámce (Data Frame) pandas. Poté vytvoří slovník (Dictionary) pro uložení jedinečných názvů sloupců a jim odpovídajících dat tak, že prochází jednotlivé sloupce v datovém rámci. Skript spojí kódy využití půdy v abecedním pořadí, aby vytvořil klíč pro slovník, a odpovídajícím způsobem připojí nebo vytvoří dvojice klíč-hodnota před vytvořením nového datového rámce a jeho exportem do nového souboru aplikace Excel s názvem "output.xlsx".

```
import pandas as pd
# Read the Excel file into a pandas dataframe
df = pd.read_excel('input.xlsx')
# Create a dictionary to store the unique column names and their data
columns_dict = {}
# Loop through each column in the dataframe
for col in df.columns:
    # Get the unique land use codes in the column
    land_uses = sorted(col.split())
        # Create a key for the dictionary by concatenating the land use codes
in alphabetical order
    key = '_'.join(land_uses)
        # If the key already exists in the dictionary, append the column data
to the existing value
    if key in columns_dict:
        columns_dict[key] += df[col]
    # Otherwise, create a new key-value pair in the dictionary
    else:
        columns_dict[key] = df[col]
# Create a new dataframe from the dictionary
new_df = pd.DataFrame(columns_dict)
# Add the first column from the original dataframe to the new dataframe
new_df[df.columns[0]] = df[df.columns[0]]
# Save the new dataframe to a new Excel file
new_df.to_excel('output.xlsx', index=False)
```

Výsledný soubor MS Excel obsahuje pro 100 vybraných evropských měst **3460 unikáních frekventovaných sad sousednosti využití území do 100 metrů** je součástí příloh (přílohy: Vystupni_Data/MSExcel_soubory/PripadovaStudie2_mestaEVROPA_sumarizacni_matice.xlxs).

Lze přijmout předpoklad, že podobná města budou mít podobnou podporu stejných frekventovaných sad. Obdobně lze hledat nepodobnosti za předpokladu, že nepodobná města budou mít rozdílnou podporu stejné frekventované sady, popřípadě že se frekventovaná sada nebude u srovnávaného města vůbec nacházet.

Obecně lze říci, že **čím větší je vzdálenost bufferu u analýzy sousednosti v rámci ArcGIS Pro, tím delší je mediánový počet prvků v jedné transakci**. Při experimentování s různými vzdálenosti bylo zjištěno, že např. u vzdálenosti 400 m je počet unikátních frekventovaných sad u 100 vybraných měst roven přibližně 180 000. Zde již narážíme na to, že klesá vypovídající hodnota použité metody. Lze říci, že při nastavení vyšší vzdálenosti, např. zmíněných 400 m sousedí "vše se vším" a z transakcí se vytrácí lokální vzory, které jsou klíčové pro následnou analýzu. Dále vzniká velké množství unikátních sad, které mají zastoupení pouze v malém počtu měst a u většiny datasetu nabývají podpory 0 %.

4.7 Příprava dat pro případovou studii – česká města

V rámci druhé případové studie byla zkoumána všechna česká města z datové sady Urban Atlas 2018. Pro všechna města byla stažena zdrojová data, pomocí nástroje v prostředí ArcGIS Pro byla vypočítána data sousednosti s parametrem vzdálenosti 100 m. Výstupem je tedy 15 textových souborů reprezentující sousednosti. Následně byla tato data spojena do jednoho souboru pomocí realizovaného scriptu JoinTXT_Files.ipynb (příloha: Vystupni_Data/SCRIPTY/JoinTXT_Files/JoinTXT_Files.ipynb).

Výstupem scriptu je jeden textový soubor s názvem cities_merged.txt (příloha: Vystupni_Data/FREKVENTOVANE_SADY/FS_Czechia_merged/cities_merged.txt), který obsahuje všechna transakční data všech 15 měst dohromady. Takto připravený soubor je dále použit pro výpočet frekventovaných sad pomocí Jupyter Notebook s minimální podporou 5 %. Vypočítaná data byla dále zkoumána ve snaze nalézt společné rysy českých měst a výsledky jsou prezentovány v kapitole 5.2 Případová studie – česká města.

```
#import necessary packages
from ntpath import basename
import os, glob, shutil
import importlib
# This function is used to list all subfolders in a folder
def list_folders(folder_path):
    return glob.glob(os.path.join(folder_path, '*'))
# This function is used to list all files of a specific type in a folder
def list_files(folder_path,file_type):
    return glob.glob("{}/*.{}".format(folder_path,file_type))
# This function is used to make a new folder/directory
# Note that if folder_path exists then remove this folder and its subfolders
def makedirs(folder_path):
    if not os.path.exists(folder_path):
        os.makedirs(folder_path)
```

```
    else:
        shutil.rmtree(folder_path)  # Removes all the subdirectories!
        os.makedirs(folder_path)
f = open("data/cities_merged.txt","w")
print("Open")
counter=0
files = list_files("data\\input_data","txt")
print(files)  # list of all files in directory
print(len(files)) # count of files
# for all files in directory All\\TXT
for j in range(len(files)):
    input_file = files[j]
    print(input_file)
    g = open(input_file,"r")
    row = g.readline()
    while row !="":
        row = g.readline()
        f.write(row)
        counter=counter+1
print(counter)
```

5 VÝSLEDKY

5.1 Popis vybraných měst

V rámci této kapitoly je popsán způsob, jak lze interpretovat výsledky vypočítaných frekventovaných sad sousednosti využití území. Tento krok je předstupněm k analýze podobnosti v rámci kapitoly 5.3 Případová studie – evropská města a 5.4 Porovnání s výsledky studie (Dobesova 2020). Jsou blíže popsána 2 vybraná města a to: Cheltenham (Spojené království) a Prešov (Slovensko). Pro každé město je vytvořena příloha obsahující vypočítané frekventované sady s minimální podporou 5 %. Následně jsou identifikovány potencionální specifické frekventované sady v rámci měst. Ty jsou následně popsány.

5.1.1 Cheltenham (Spojené království)

Cheltenham je lázeňské město v jihozápadní Anglii, v hrabství Gloucestershire. Město se nachází na úpatí Cotswold Hills a je známé svou architekturou a elegantními zahradami. Město má přibližně 116 000 obyvatel a díky malebnému okolí a kulturním památkám je oblíbeným cílem turistů. Vyznačuje se kombinací obytných, komerčních a průmyslových ploch. Město má dobře rozvinuté centrum s řadou obchodů, restaurací a zábavních podniků. V okolí městské části se nacházejí také rozsáhlé venkovské oblasti, včetně zemědělské půdy a volných zelených ploch.

Frekventované sady byly vypočítány na základě transakčních dat sousednosti získaných aplikací nástroje SearchDistinctLanduse_SpatialJoin při nastavení vzdálenosti bufferu 100 m. Minimální podpora pro frekventované sady byla stanovena na 5 %. Bylo vypočítáno celkem **277 frekventovaných sad**. Výčet sad do podpory 20 % lze vidět v (Tabulka 7). Kompletní výčet frekventovaných sad do podpory 5 % je součástí příloh (příloha: Vystupni_Data/MSExcel_soubory/PripadovaStudie2_mestaEVROPA.xlsx) na listu CHELTENHAM.

V rámci Cheltenhamu jsou nejfrekventovanější sady sousednosti obsahující **průmyslové, komerční a veřejné plochy (12100) s nesouvislou hustou městskou zástavbou (11210)** 12100 11210 s podporou 60 % (Obrázek 27). Takováto místa můžeme vidět ve středu města, kde je nesouvislá hustá zástavba (tmavě červená) doplněna o pravděpodobně komerční plochy (fialová). Pro město je typická kombinace **průmyslové, komerční a veřejné zástavby (12100), nesouvislé husté městské zástavby (11210) a nesouvislé středně husté městské zástavby (11220)** 12100 11210 11220 s podporou 45 %. Vybrané frekventované sady sousedství lze vidět v (Tabulka 6).

Obrázek 27 Typická středně hustá zástavba (11220) v Cheltenhamu (Spojené království) – (zdroj: https://cs.wikipedia.org/wiki/Cheltenham_%28Spojen%C3%A9_kr%C3%A1lovstv%C3%AD%29#/media/Soubor:Cheltenham.from.leckhampton.arp.jpg).

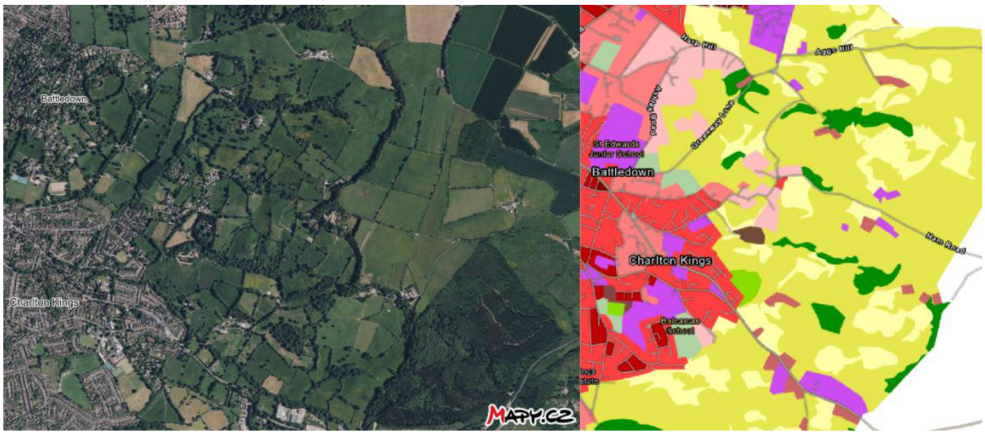
Tabulka 6 Vybrané FS pro město Cheltenham (Spojené království).

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3
60	714	12100	11210	
56	663	11210	11220	
54	644	12100	11220	
45	539	12100	11210	11220
28	331	11220	11230	
28	332	11210	14100	
27	322	11220	14100	
24	289	23000	21000	
24	288	12100	11230	
23	275	11210	11220	14100
23	277	12100	11210	14100

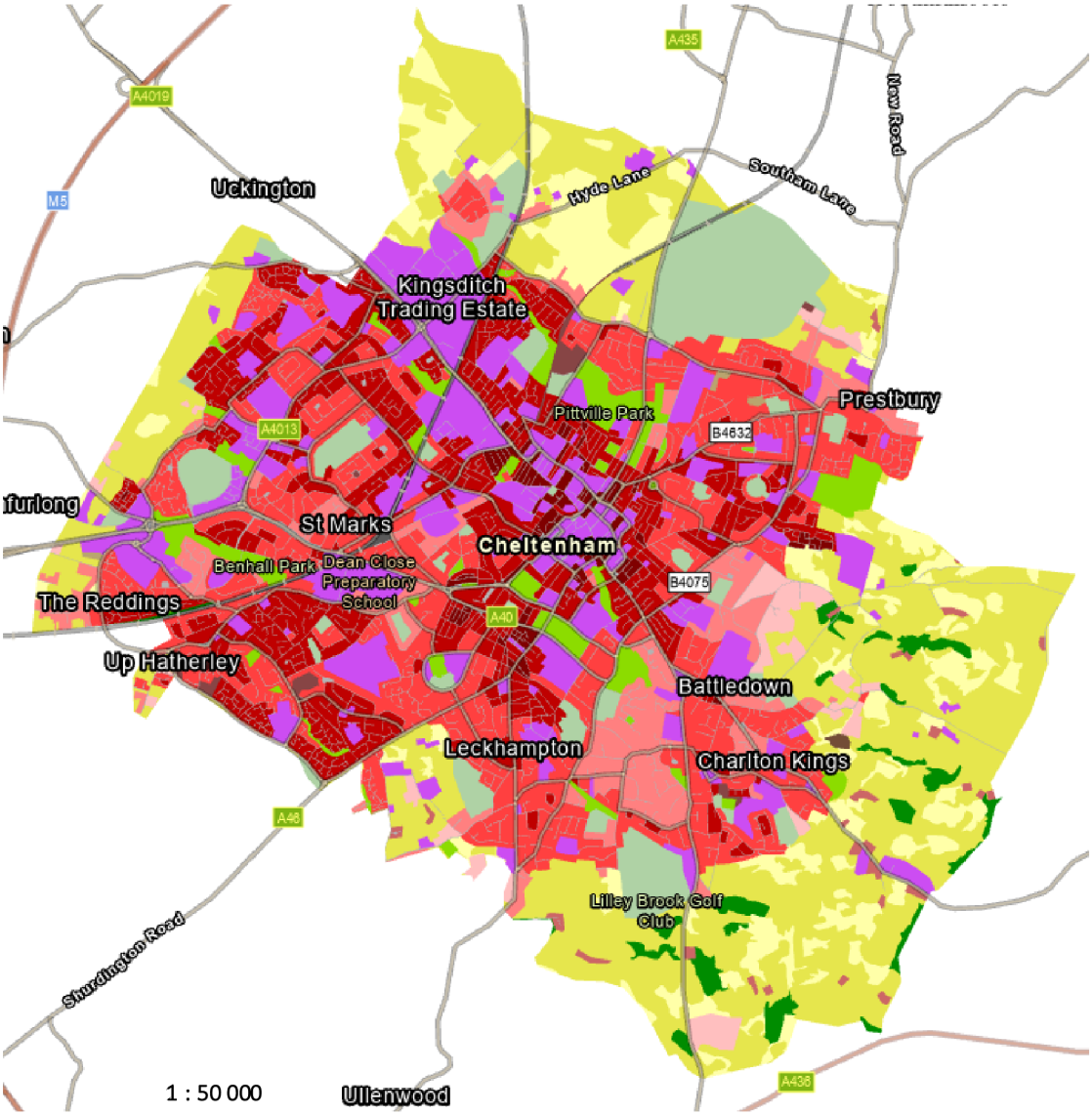
Rovněž poměrně častou je frekventovaná sada **nesouvislé středně husté zástavby (11220)** s **nesouvislou málo hustou městskou zástavbou (11230)** 11220 11230 s podporou 28 %. Tento druh sousedství můžeme nalézt například v části města Charlton Kings, Prestbury nebo Leckhampton (Obrázek 28).

Obrázek 28 Ukázka nesouvislé málo husté městské zástavby (12230) v části Cheltenhamu (Charlton Kings) – (zdroj: https://en.wikipedia.org/wiki/Charlton_Kings#/media/File:Charlton_Kings.jpg).

Co se týče okolí města, nejfrekventovanějším sousedstvím v okolí s podporou 24 % je kombinace 23000 21000 **pastviny (23000) a orné půdy (21000)**. Tento druh sousedství lze vidět ve východní části vymezeného území (Obrázek 29). Malé nerovnoměrně rozmístěné oblasti orné půdy jsou obklopeny pastvinami a také malými lesy.



Obrázek 29 FS – orná půda (21000) a pastviny (23000) - při východní hranici vymezené území Cheltenhamu – (zdroj: mapy.cz).



Obrázek 30 Cheltenham (Spojené království) - UA 2018.

Tabulka 7 Frekventované sady města Cheltenham (Spojené království) - do podpory 20 %.

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3	Landuse4
95	1129	12220			
75	887	12100			
74	880	12220	12100		
72	857	11210			
72	856	12220	11210		
69	815	11220			
69	813	12220	11220		
60	714	12100	11210		
60	713	12220	12100	11210	
56	663	11210	11220		
56	663	12220	11210	11220	
54	644	12100	11220		
54	643	12220	12100	11220	
45	539	12100	11210	11220	

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3	Landuse4
45	539	12220	12100	11210	11220
34	404	23000			
33	387	14100			
32	379	11230			
32	378	12220	11230		
32	383	12220	14100		
29	345	14200			
29	343	12220	14200		
29	348	12220	23000		
28	331	11220	11230		
28	331	12220	11220	11230	
28	332	11210	14100		
28	331	12220	11210	14100	
27	322	11220	14100		
27	321	12220	11220	14100	
27	320	12100	14100		
27	318	12220	12100	14100	
25	295	11100			
25	295	12220	11100		
25	296	21000			
24	287	11210	11100		
24	287	12220	11210	11100	
24	289	23000	21000		
24	288	12100	11230		
24	287	12220	12100	11230	
23	275	11220	14200		
23	275	12220	11220	14200	
23	267	11210	14200		
23	267	12220	11210	14200	
23	274	12100	14200		
23	273	12220	12100	14200	
23	275	11210	11220	14100	
23	275	12220	11210	11220	14100
23	277	12100	11210	14100	
23	276	12220	12100	11210	14100
22	255	12100	11100		
22	255	12220	12100	11100	
22	256	12100	11220	11230	
22	256	12220	12100	11220	11230
22	263	12100	11220	14100	
22	262	12220	12100	11220	14100
21	248	12100	11210	11100	
21	248	12220	12100	11210	11100
21	245	12220	21000		
21	243	11210	11230		
21	243	12220	11210	11230	
20	238	12220	23000	21000	

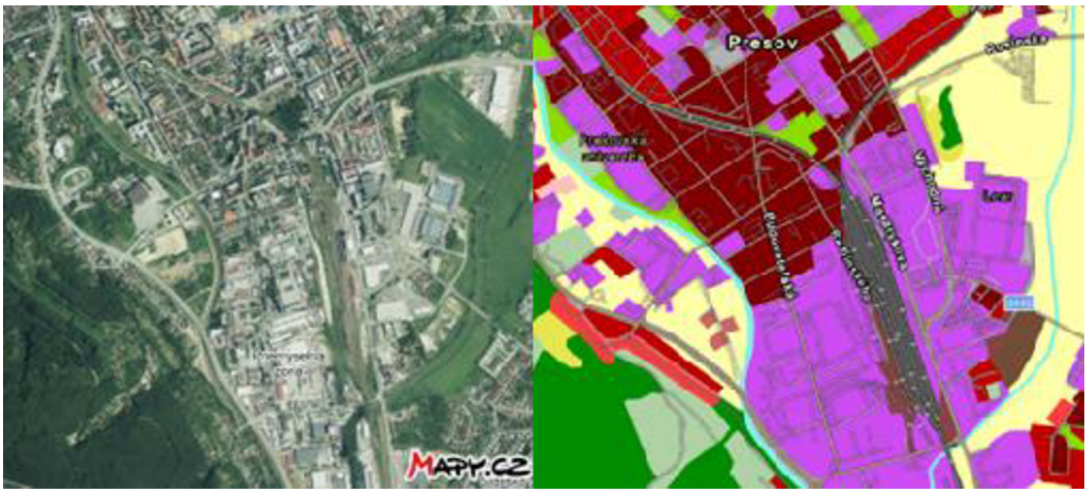
5.1.2 Prešov (Slovensko)

Prešov se nachází na východě Slovenska v Prešovském kraji. Město leží v údolí řeky Torysa, obklopené východními Karpaty. Prešov má přibližně 90 000 obyvatel a je třetím největším městem na Slovensku. Prešov se vyznačuje kombinací obytných, obchodních a průmyslových ploch. V centru města se nacházejí historické památky, obklopené komerčními částmi, jako jsou obchody, restaurace a kavárny. Průmyslové oblasti jsou soustředěny především v jižní části města, kde se nacházejí různé výrobní podniky a podniky služeb, včetně automobilového průmyslu. Prešov má dobře rozvinutou dopravní síť, včetně hlavních silnic, železnic a veřejné dopravy, která ho spojuje s ostatními částmi Slovenska a sousedními zeměmi.

Kompletní výčet frekventovaných sad do podpory 5 % je součástí (příloha: Vystupni_Data/MSExcel_soubory/PripadovaStudie2_mestaEVROPA.xlsx) v záložce PRESOV. Pro město Prešov bylo při minimální podpoře 5 % vypočítáno **455 frekventovaných sad**. Pro Prešov je nejfrekventovanější sadou sousedství **průmyslových, komerčních a veřejných ploch (12100) a souvislé husté městské zástavby (11100) s podporou 51 %**. Z této dvouprvkové frekventované sady vznikla také tříprvková frekventovaná sada doplněná o nesouvislou hustou městkou zástavbu (11210) s podporou 40 %. Vybrané frekventované sady sousedství lze vidět v (Tabulka 8).

Tabulka 8 Vybrané FS pro město Prešov (Slovensko).

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3	Landuse4	Landuse5
51	671	12100	11100			
49	637	11210	11100			
40	522	11210	21000			
40	524	12100	21000			
40	519	12100	11210	11100		
35	463	12100	11100	14100		
35	452	11210	14100			
34	448	11210	14200			
28	361	12220	12100	11210	11100	14100
22	285	21000	31000			
22	283	14100	14200			



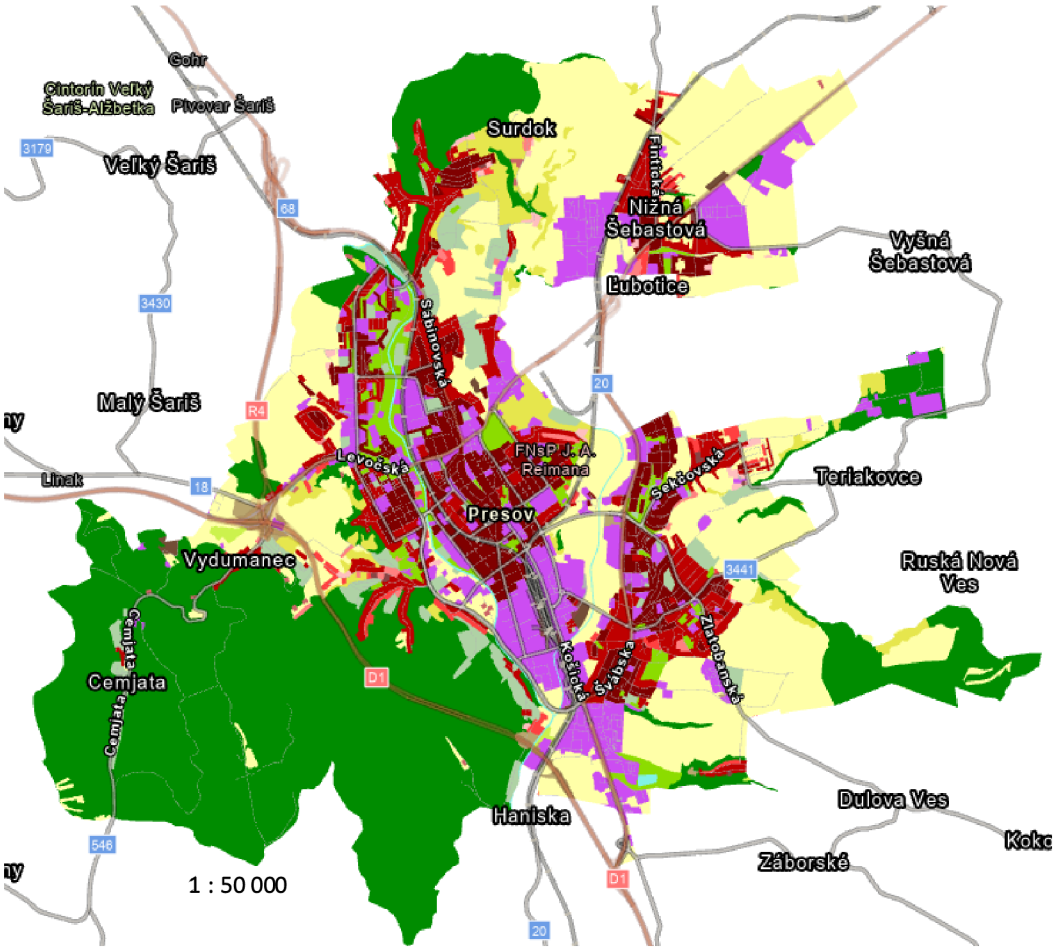
Obrázek 31 Kombinace souvislé husté městské zástavby (11100) a průmyslové, komerční a veřejné zástavby (12100) v Prešově – (zdroj: mapy.cz).

Poměrně časté jsou rovněž **sousednosti s ornou půdou (21000)**, která opklopuje město ze všech stran kromě severozápadu, kde se nachází souvislý les. S sousedství s ornou půdou se vyskytuje **nesouvislá hustá městaká zástravba (11210)** s podporou 40 %. Takovýto typ sousedsví lze vidět v městské části Solivar, kde nesouvislá zástavba přechází v ornou půdu (Obrázek 32). Dále se orná půda nachází v sousedství s **průmyslovými, komerčními a veřejnými plochami (12100)** s podporou 40% (Obrázek 31). Souvislá hustá městská zástavba (11100), která se nachází v centru města, s ornou půdou nesousedí.

Obrázek 32 Kombinace orné půdy (21000) a nesouvislé husté městské zástavby (11210) v okrajové části Prešova Solivar – (zdroj: https://domalenka.cz/atrakce/solivar-pri-presov).

Zajímavou čtyřkombinací je sousednost **průmyslových, komerčních a veřejných ploch (12100), nesouvislé husté městské zástavby (11210), souvislé husté městské zástavby (11100) a městské zeleně (14100)**. Jedná se o rozšíření tříprvkové sady s podporou 40 % o městskou zeleň. Sada má podporu 28 %. Městská zeleň se nachází převážně ve středu města a také v severní části města podél řeky Torysy. Městská zeleň se rovněž vyskytuje v sousedství v kombinaci se **sportovišti (14200)** s podporou 22 %.

Obrázek 33 Kombinace městské zeleně (14100) se souvislou hustou městskou zástavbou (11100), nesouvislou hustou městskou zástavbou a průmyslovými, komerčními a veřejnými plochami (12100) – (zdroj: https://sk.wikipedia.org/wiki/Torysa_%28rieka%29#/media/S%C3%BAbor:Slovakia_Town_Presov_Pc333.jpg)



Obrázek 34 Prešov (Slovensko) - UA 2018.

Tabulka 9 Frekventované sady města Prešov (Slovensko) - do podpory 20 %.

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3
97	1266	12220		
73	948	12100		
72	941	12220	12100	
69	898	11210		
69	898	12220	11210	
61	802	11100		
61	802	12220	11100	
60	789	21000		
58	755	12220	21000	
51	671	12100	11100	
51	671	12220	12100	11100
49	637	11210	11100	
49	637	12220	11210	11100
49	634	12100	11210	
49	634	12220	12100	11210
44	579	14100		
44	578	12220	14100	
41	541	14200		
41	538	12220	14200	
40	522	11210	21000	
40	522	12220	11210	21000

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3	Landuse4	Landuse5
40	524	12100	21000			
40	519	12220	12100	21000		
40	519	12100	11210	11100		
40	519	12220	12100	11210	11100	
39	515	11100	14100			
39	515	12220	11100	14100		
39	505	12100	14100			
39	504	12220	12100	14100		
35	463	12100	11100	14100		
35	463	12220	12100	11100	14100	
35	452	11210	14100			
35	452	12220	11210	14100		
34	448	11210	14200			
34	448	12220	11210	14200		
31	408	11210	11100	14100		
31	408	12220	11210	11100	14100	
30	385	12100	11210	14100		
30	385	12220	12100	11210	14100	
29	382	31000				
29	384	12100	14200			
29	384	12220	12100	14200		
29	376	11100	21000			
29	376	12220	11100	21000		
28	361	12100	11210	11100	14100	
28	361	12220	12100	11210	11100	14100
27	354	12220	31000			
26	333	11100	14200			
26	333	12220	11100	14200		
26	335	12100	11210	21000		
26	335	12220	12100	11210	21000	
25	323	21000	14200			
25	322	12220	21000	14200		
24	311	11220				
24	310	12220	11220			
24	310	12100	11210	14200		
24	310	12220	12100	11210	14200	
24	311	11210	11100	21000		
24	311	12220	11210	11100	21000	
23	298	12100	11100	21000		
23	298	12220	12100	11100	21000	
22	285	21000	31000			
22	283	14100	14200			
22	283	12220	14100	14200		
22	287	11210	11100	14200		
22	287	12220	11210	11100	14200	
21	278	11210	11220			
21	278	12220	11210	11220		
21	276	12100	11100	14200		
21	276	12220	12100	11100	14200	
20	261	12220	21000	31000		
20	262	11210	31000			
20	262	12220	11210	31000		
20	267	11210	21000	14200		
20	267	12220	11210	21000	14200	

V Cheltenhamu jsou nejfrekventovanější sady sousednosti obsahující **průmyslové, komerční a veřejné plochy (12100)** s **nesouvislou hustou městskou zástavbou (11210)**. Rovněž často se vyskytují kombinace s **nesouvislou středně hustou městskou zástavbou (11220)** a **nesouvislou málo hustou městskou zástavbou (12230).** Okolí města je vyznačuje kombinací orné půdy (21000) a pastvin (23000).

Naopak pro Prešov je poměrně typická kombinace sousedství **průmyslových, komerčních a veřejných ploch (12100) a souvislé husté městské zástavby (11100)**, která u Cheltenhamu není frekventovaná**.** Sousedství s ornou půdou (21000), které obklopují město ze všech stran s výjimkou severozápadu, kde se rozkládá souvislý les, jsou také poměrně častá.

Celkem překvapivě jsou sousedství se zelení frekventovanější v Prešově, které je průmyslovým městem oproti lázeňskému městu Cheltenham. Na druhou stranu je Cheltenham obecně charakterističtější řidším typem zástavby (čtvrti rodinných domů), které pravděpodobně mají vlastní zahrady. Celkově lze říci, že se jedná o dvě velice různá města. Podstatnou rozdílnost naznačují **značně rozdílné frekventované sady využití území**. Lze tedy do určité míry říci, že **frekventované sady sousedství charakterizují uspořádání města**.

5.2 Případová studie – česká města

Cílem případové studie bylo **souhrnně popsat česká města v rámci datové sady UA 2018**. Tato datová sada obsahuje celkem 15 českých měst a jsou jimi: Praha, Brno, Ostrava, Plzeň, Ústí nad Labem, Olomouc, Liberec, České Budějovice, Hradec Králové, Pardubice, Zlín, Karlovy Vary, Jihlava, Most a Chomutov. Jedná se tedy o krajská města České republiky doplněná o Most a Chomutov.

Po vypočítání transakčních dat sousednosti jednotlivých měst **byly vypočítány frekventované sady všech měst dohromady**, čehož bylo docíleno sloučením transakčních dat jednotlivých měst do jednoho souboru, který dále vstupoval do výpočtu frekventovaných sad. Byla stanovena minimální podpora 5 %. Celkem bylo zjištěno 319 sad. Mediánová délka transakce byla 3.

Mezi sady s nejvyšší podporou patřily jednoprvkové frekventované sady (Tabulka 10): **ostatní silnice (12220) = 97 %, nesouvislá hustá zástavba (11210) = 69 %, souvislá městská zástavba (11100) = 61 %** atd. Tyto frekventované sady vznikají z podstaty metody (viz kapitola 2.3 Frekventované sady) a nevypovídají o konkrétní sousednosti. Nicméně jednoprvkové transakce říkají, že daný typ landuse existuje ve zdrojových transakcích v kombinaci s nějakým dalším typem landuse. Je to vlastně počet konkrétního landuse, který se vykytuje ve všech vstupních transakcích. Takovéto sady nesou jen základní výchozí informaci, kdy můžeme porovnat, která kategorie je v pravidlech četnější než jiná. Na příkladu tabulky 10 je četnější kategorie **nesouvislá hustá zástavba (11210) = 69 %** než **souvislá městská zástavba (11100) = 61 %.** V případě kategorií zástavby se mohou města právě lišit podporou a pořadím podle druhu zástavby. Informace z jednoprvkových sad lze považovat za méně důležitou. Obdobně méně důležité jsou pravidla s dvěma hodnotami využití území, z nichž jedna je silnice (12220, 12210). Tyto sady z pohledu sousednosti pouze reflektují vstupní data, ve kterých silnice „krájí“ městskou mozaiku na části. Rovněž není překvapující, že se v blízkosti většiny polygonů v městech nachází silnice a je to typický jev.

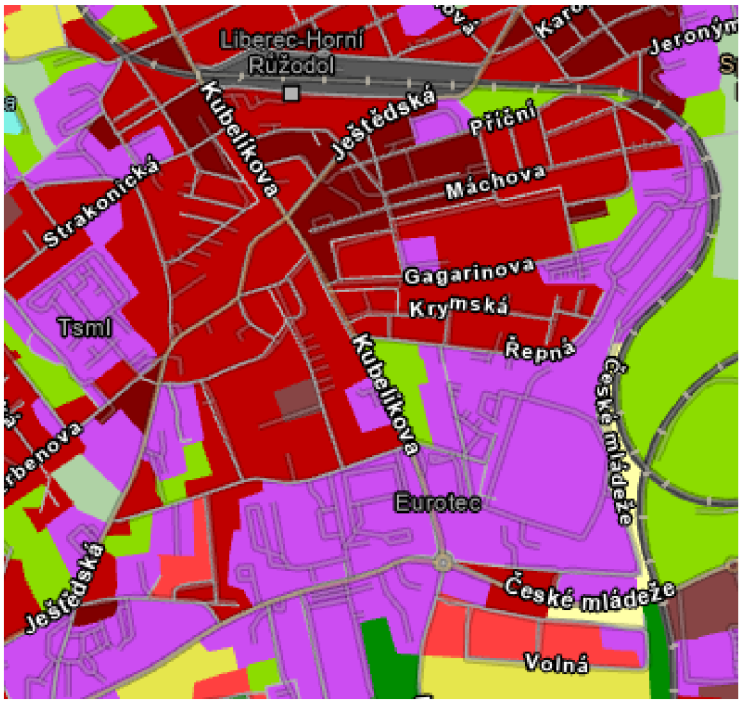
Tabulka 10 Ukázka jednoprvkových frekventovaných sad a dvouprvkových frekventovaných sad pro česká města (jeden z prvků je vždy silnice 12220).

% Podpora	Abs. Podpora	Landuse1	Landuse2
97	44975	12220	
71	32694	12100	
70	32414	12220	12100
69	31777	11210	
68	31670	12220	11210
61	28458	11100	
61	28446	12220	11100

Prvním zajímavou frekventovanou sadou je kombinace 12100 a 11210 s podporou 49 %. Jedná se o kombinaci **průmyslové, komerční a veřejné zástavby (12100) a nesouvislé husté městské zástavby (11210)**. Hlavním problémem interpretace je příliš široké rozpětí kategorie 12100. Jedná se o průmyslovou, komerční, veřejnou či vojenskou zástavbu. Výše zmíněná frekventovaná sada tedy může poukazovat na kombinaci nesouvislé husté městské zástavby s průmyslovou zónou, nákupními zónami, školami, nemocnicemi či dokonce vojenskými objekty. Je tedy potřeba zmínit, že v rámci této sady může výše zmíněný typ zástavby sousedit se všemi zmíněnými kategoriemi najednou. Na základě vstupních dat (UA 2018) není možné jednotlivé kategorie dále odlišovat. Nelze tedy vyvozovat závěry typu: hustá zástavba sousedí s komerční plochou, protože se ve skutečnosti může jednat o plochu průmyslovou, či kombinaci více typů využití území.

Tabulka 11 FS – průmyslové, komerční a veřejné plochy (12100) s nesouvislou hustou městskou zástavbou (11210).

% Podpora	Abs. Podpora	Landuse1	Landuse2
49	22703	12100	11210

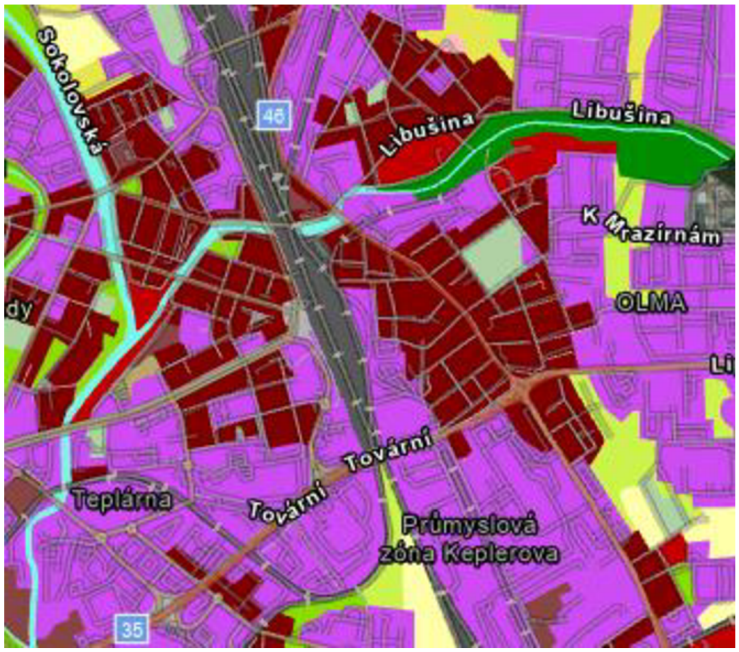


Obrázek 35 Ukázka FS – průmyslové, komerční a veřejné plochy (12100) s nesouvislou hustou městskou zástavbou (11210) – Liberec část Horní Hanychov.

Další frekventovanou sadou je kombinace 12100 a 11100 s podporou 48 %. Zde se jedná o kombinaci **průmyslové, komerční a veřejné zástavby (12100) se souvislou hustou městskou zástavbou (11100)**. Takováto kombinace indikuje nahuštěné bloky budov doplněné o pravděpodobně komerční či veřejnou zástavbu. Tento druh zástavby (souvislá hustá městská zástavba) se vyskytuje například v historických centrech měst nebo se jedná o starší hustou zástavbu původních vesnic (kolem návsí apod.), která byla připojena k větším městům.

Tabulka 12 FS – průmyslové, komerční a veřejné plochy (12100) se souvislou hustou městkou zástavbou (11100).

% Podpora	Abs. Podpora	Landuse1	Landuse2
48	22459	12100	11100

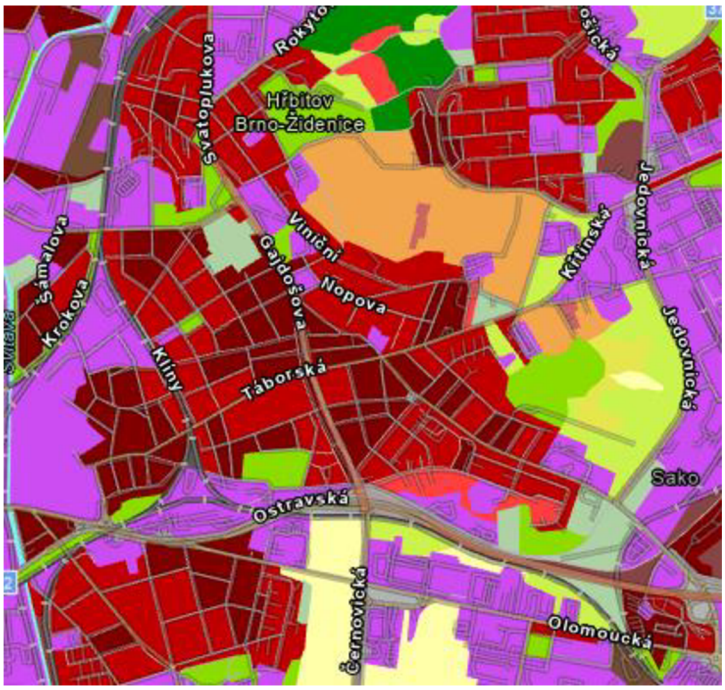


Obrázek 36 Ukázka FS – průmyslové, komerční a veřejné plochy (12100) s nesouvislou hustou městskou zástavbou (11210) - příklad Olomouc Holice.

Dalším typickým rysem českých měst je **přechod mezi souvislou městskou zástavbou (11100) a nesouvislou hustou městskou zástavbou (11210) s podporou 47 %**. Jedná se převážně o přechody mezi hustě zastavěným historickým centrem a zástavbou, která jej obklopuje. Popřípadě se jedná o okrajové části sídlišť.

Tabulka 13 FS – nesouvislá hustá městská zástavba (11210) se souvislou hustou městskou zástavbou (11100).

% Podpora	Abs. Podpora	Landuse1	Landuse2
47	21981	11210	11100



Obrázek 37 Ukázka FS – nesouvislá hustá městská zástavba (11210) se souvislou hustou městskou zástavbou (11100) - příklad Brno Židenice.

Další skupinou využití území, která je v rámci vybraných český měst **typická, je městská zeleň**. Zeleň se nejčastěji nachází v sousedství silnic (podpora 48 %) a **průmyslových, komerčních a veřejných ploch (podpora 40 %)**. Pokud bychom chtěli zkoumat sousednost zeleně a zástavby, nejčetnější je **kombinace se souvislou hustou městskou zástavbou (11100) s podporou 37 %**. Druhou nejčetnější zástavbou v **kombinaci se zelení je nesouvislá hustá městská zástavba (11210) s podporou 35 %.** Zajímavou sadou s pěti využitími území je kombinace silnice (12220), průmyslové, komerční a veřejné zástavby (12100), nesouvislé husté městské zástavby (11210), souvislé husté městské zástavby (11100) a městské zeleně (14100) s podporou 23 % (poslední řádek Tabulka 12). Městská zeleň se tedy většinou nachází v blízkosti dvou nejsytěji červených zástaveb. Tato sada poukazuje na určitou pestrost využití území v rámci českých měst.

Tabulka 14 FS využití území – městská zeleň (14100).

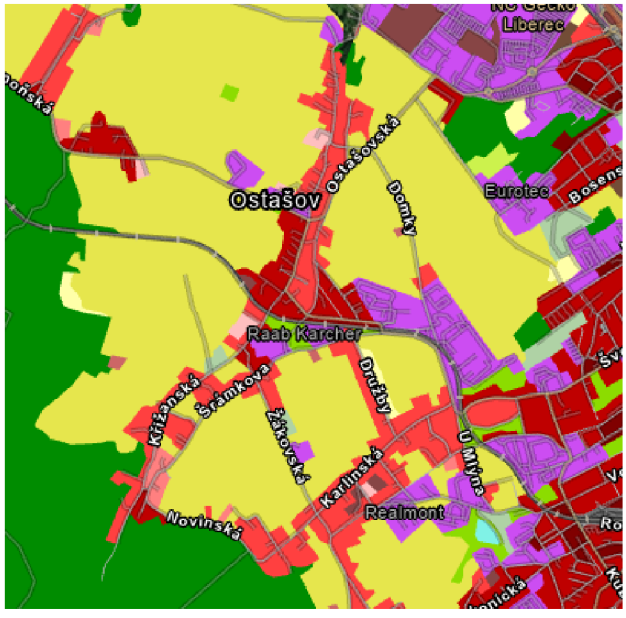
% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3	Landuse4	Landuse5
48	22182	14100				
48	22116	12220	14100			
40	18410	12100	14100			
37	17176	11100	14100			
35	16431	11210	14100			
23	10584	12220	12100	11210	11100	14100

Obrázek 38 Ukázka FS – souvislá hustá městská zástavba (11100) s městskou zelení (14100) a průmyslové, komerční a veřejné plochy (12100) s městkou zelení (14100) – příklad Komenského sady Ostrava a Smetanovy sady Olomouc.

Stále poměrně často zastoupeným typem zástavby je **nesouvislá středně hustá městská zástavba (11220)**. Tento typ zástavby se nejčastěji nachází v sousedství v kombinaci nesouvislou hustou městskou zástavbou (11210) podporou 18 %. Dále pak v kombinaci s průmyslovými, komerčními a veřejnými plochami (13 %) a v kombinaci s pastvinami (13 %), ornou půdou (21000) s podporou 12 %. Jedná se o předměstí.

Tabulka 15 FS využití území – nesouvislá středně hustá zástavba (11220).

% Podpora	Abs. Podpora	Landuse1	Landuse2
18	8327	11210	11220
13	6027	12100	11220
12	5428	23000	11220
12	5360	21000	11220



Obrázek 39 Ukázka FS – nesouvislá středně hustá zástavba (11220) s pastvinami (23000) – příklad Liberec Ostašov a okolí.

U českých měst nejsou nejfrekventovanější sady obsahující dvě nejřidší skupiny zástavby – **nesouvislou městskou zástavbu nízké hustoty (11230)** a **nesouvislou městskou zástavbu velmi nízké hustoty (11240)**. Podpora pravidel obsahující tyto typy zástavby má podporu okolo 7 %. Pod těmito typy zástavby si můžeme představit čtvrtě rodinných domů s velkými zahradami a rozestupy mezi jednotlivými domy. V rámci vybraných českých měst nejsou výrazně zastoupeny. Naopak velmi typické jsou například pro britská města (kapitola 5.3).

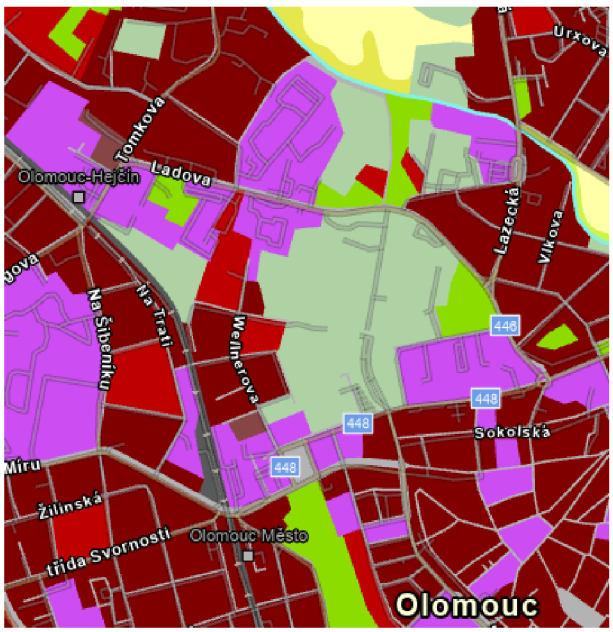
Tabulka 16 FS využití území – nesouvislá městská zástavba nízké (11230) a velmi nízké hustoty (11240).

% Podpora	Abs. Podpora	Landuse1	Landuse2
6	2756	11210	11240
5	2455	11210	11230

V rámci měst se kromě komerční, veřejné a bytové zástavby (červené kategorie) nachází rovněž **sportoviště (14200)**. Sportoviště se nejčastěji nacházejí v sousedství dohromady s průmyslovými, komerčními a veřejnými plochami (12100) s podporou 24 %. Dále pak v kombinaci s nejhustšími typy zástavby – se souvislou hustou městkou zástavbou (11100) s podporou 21 % a s nesouvislou hustou městskou zástavbou (11210) s podporou 23 %. Zajímavou kombinací je rovněž čtveřice využití území: souvislá hustá městská zástavba (11100), nesouvislá hustá městská zástavba (11210) a sportoviště (14200) samozřejmě v kombinaci se silnicí. Tato sada indikuje sousedství dvou typů zástavby se sportovištěm. Jedná se pravděpodobně o veřejná sportoviště mezi jednotlivými čtvrtěmi.

Tabulka 17 FS využití území – sportoviště (14200).

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3	Landuse4	Landuse5
24	11119	12100	14200			
23	10803	11210	14200			
21	9821	11100	14200			
17	7990	12220	11210	11100	14200	



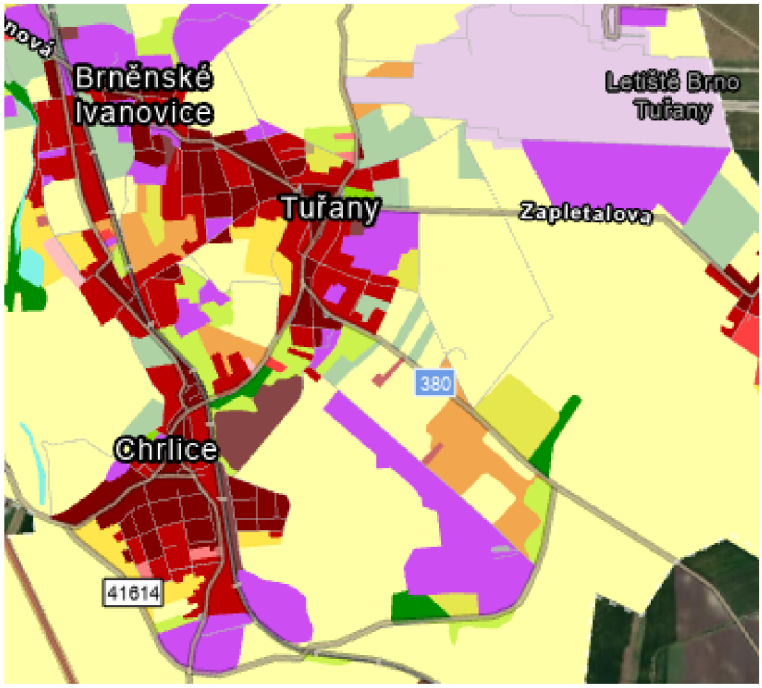
Obrázek 40 FS sportoviště (14200) v kombinaci s průmyslovými, komerčními a veřejnými plochami (12100) – Olomouc okolí fotbalového stadionu.

Z rurálních kategorií využití území na území urbánních jader vybraných měst mají největší podporu pravidla obsahující **lesy (31000), ornou půdu s jednoletými rostlinami (21000) a pastviny (23000)**. Naopak nejsou zastoupeny sady (25000), půda s trvalými plodinami (22000) a komplexní půda se smíšeným typem pěstování (24000). Orná půda se nachází v sousedství v kombinaci s nesouvislou hustou městkou zástavbou (11210) s podporou sady 24 %. Dále se nachází v sousedství v kombinaci s průmyslovými, komerčními a veřejnými plochami (12100) s podporou 22 % napříč vybranými městy.

To dokladuje trend posledních 20 let, kdy se města rozrůstají na okrajích na úkor původní orné půdy a vznikají zde logistická centra a průmyslové areály. Méně časté je v současnosti rozrůstání měst formou suburbanizace řídkou zástavbou rodinných domků.

Tabulka 18 FS využití území – orná půda (21000).

% Podpora	Abs. Podpora	Landuse1	Landuse2
24	11206	11210	21000
22	10101	12100	21000



Obrázek 41 Ukázka FS – využití území orná půda (21000) v kombinaci s průmyslovými, komerčními a veřejnými plochami (12100) – příklad Brněnské Ivanovice.

Celkově byly pro vybraná česká města nalezeny vypovídající frekventované sady. Jejich výčet při minimální podpoře 20 % lze vidět v (Tabulka 19). Celá tabulka při podpoře 5 % je součástí příloh (příloha: Vystupni_Data/MSExcel_soubory/PripadovaStudie1_mestaCR.xlsx). Výše zmíněné kombinace využití území jsou pro česká města typické. Pokud by byly provedeny obdobné studie pro jednotlivé národní státy v rámci datasetu, bylo by možné jednotlivé státy porovnat a nalézt podobnosti, popřípadě rozdílnosti. V rámci aplikované metodiky je možno změnit minimální podporu FS, velikost bufferu pro analýzu sousednosti a samotnou oblast zájmu. Namísto urbánních jader by mohlo býti použito FUA nebo administrativní jednotky.

Tabulka 19 Frekventované sady všech českých měst v rámci datasetu UA 2018, do podpory 20 %.

% Podpora	Abs. Podpora	Landuse1	Landuse2	Landuse3	Landuse4	Landuse5
97	44975	12220				
71	32694	12100				
70	32414	12220	12100			
69	31777	11210				
68	31670	12220	11210			
61	28458	11100				
61	28446	12220	11100			
49	22703	12100	11210			
49	22677	12220	12100	11210		
48	22182	14100				
48	22116	12220	14100			
48	22459	12100	11100			
48	22453	12220	12100	11100		
47	21981	11210	11100			
47	21976	12220	11210	11100		
40	18410	12100	14100			
40	18366	12220	12100	14100		
37	17051	21000				
37	17176	11100	14100			
37	17171	12220	11100	14100		
36	16715	12100	11210	11100		
36	16713	12220	12100	11210	11100	
35	16123	12220	21000			
35	16431	11210	14100			
35	16412	12220	11210	14100		
32	14640	31000				
32	14828	23000				
31	14333	14200				
31	14197	12220	14200			
31	14560	12100	11100	14100		
31	14556	12220	12100	11100	14100	
30	13957	12220	23000			
29	13625	12220	31000			
29	13290	12100	11210	14100		
29	13280	12220	12100	11210	14100	
28	12802	11210	11100	14100		
28	12799	12220	11210	11100	14100	
24	11119	12100	14200			
24	11079	12220	12100	14200		
24	11206	11210	21000			
24	11139	12220	11210	21000		
23	10803	11210	14200			
23	10787	12220	11210	14200		
23	10586	12100	11210	11100	14100	
23	10584	12220	12100	11210	11100	14100
22	10163	11220				
22	10052	12220	11220			
22	10101	12100	21000			
21	9821	11100	14200			
21	9817	12220	11100	14200		
21	9684	11210	23000			
21	9624	12220	11210	23000		
21	9935	12220	12100	21000		
20	9131	11210	31000			
20	9063	12220	11210	31000		
20	9091	21000	23000			

5.3 Případová studie – evropská města

Cílem této případové studie je **nalezení skupin v rámci vybraných evropských měst (100)**. Metodika výběru byla přiblížena v rámci kapitoly 4.1 Výběr dat. Výběr se skládá z části z dvojic, jejichž podobnost byla zjištěna na základě obrazové podobnosti (Dobesova 2020) doplněné podle metodiky představené v kapitole 4.1 Výběr dat.

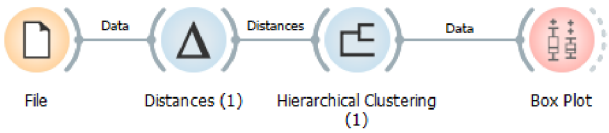
Hlavní myšlenkou pro hledání podobnosti je, že **podobná města budou mít podobnou procentuální podporu stejných frekventovaných sad sousednosti využití území**. Abychom mohli tyto procentuální shody hledat, bylo potřeba vytvořit program, který by sumarizoval výsledky jednotlivých měst. V rámci spolupráce s Univerzitou Kyoto byla vyvinuta sumarizační funkce v rámci již existujícího Jupyter Notebooku. Tato funkce, která vytváří sumarizační matici byla blíže představena v rámci kapitoly 4.6 Příprava dat pro případovou studii – evropská města. Rovněž byla představena struktura sumarizační matice. Pro připomenutí, sloupce matice tvoří unikátní (neopakující se) frekventované sady napříč všemi 100 městy. Řádky tvoří jednotlivá města a hodnoty matice tvoří procentuální podpory jednotlivých frekventovaných sad v daných městech. Pro vybraná města studie byla vypočítána **sumarizační matice**, jejíž náhled lze vidět v (Tabulka 20), celá sumarizační matice je součástí příloh (příloha: Vystupni_Data/MSExcel_soubory/PripadovaStudie2_mestaEVROPA_sumarizacni_matice.xlsx).

Tabulka 20 Náhled (prvních 20 měst a 6 frekventovaných jednoprvkových sad) sumarizační matice – procentuální podpora všech unikátních frekventovaných sad při minimální podpoře 5 % ve vybraných evropských městech.

City	11100	11210	11220	11230	11240	11300
SHKODER	34	56	30	48	16	6
SALZBURG	22	74	67	33	0	13
INNSBRUCK	22	59	62	32	8	19
MOSTAR	27	79	48	35	16	0
KORTRIJK	26	48	55	47	21	24
OOSTENDE	63	58	34	13	7	11
PLOVDIV	69	50	12	0	8	6
BASEL	16	70	68	36	8	0
BERN	6	46	67	60	22	23
LEMESOS	58	70	53	36	22	11
PLZEN	65	68	11	0	0	6
USTI_NAD_LABEM	52	78	28	10	13	9
OLOMOUC	76	52	9	0	8	7
LIBEREC	45	79	50	9	14	14
CESKE_BUDEJOVICE	68	64	26	6	0	0
HRADEC_KRALOVE	50	70	24	0	9	13
PARDUBICE	48	72	29	0	10	0
KARLOVY_VARY	41	69	42	15	5	10
AUGSBURG	69	66	19	8	0	0
NEUMUNSTER	17	42	70	21	12	9

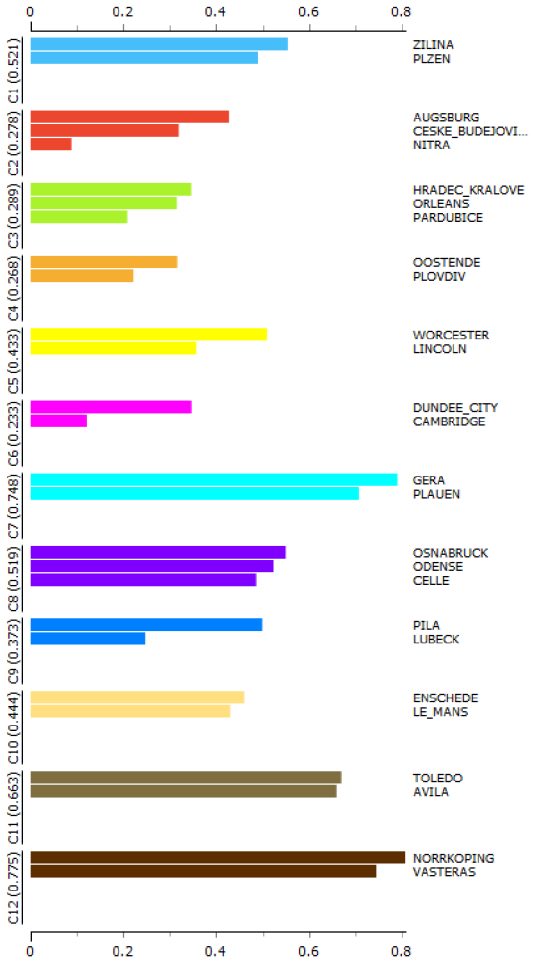
Nad touto maticí bylo provedeno hierarchické shlukování v programu Orange (Obrázek 42) s cílem nalézt skupiny (klastry) měst s podobnými frekventovanými sadami sousednosti. Byla aplikována následující nastavení. Byla vypočítána **euklidovská vzdálenost**. V **dendrogramu hierarchického shlukování** představuje vzdálenost v horní části dendrogramu vzdálenost nebo rozdílnost mezi nejodlišnějšími shluky v souboru dat. Tato vzdálenost se obvykle měří pomocí metriky vzdálenosti. Jako mezní hodnota byla stanovena vzdálenost 125. Výška dendrogramu je grafickým znázorněním vzdálenosti mezi

shluky. Čím delší je svislá čára, která spojuje dva shluky, tím větší je mezi nimi vzdálenost. Proto jsou si shluky, které jsou spojeny kratšími čarami, navzájem podobnější než shluky, které jsou spojeny delšími čarami. Dalším nastavením hierarchického shlukování v Orange je typ vázání (linkage). Bylo zvoleno **Wardovo spojování (Ward linkage)**. Tato metoda spojování se snaží minimalizovat rozptyl v rámci každého shluku při vytváření nových shluků. Jinými slovy se snaží vytvořit shluky, které jsou co nejhomogennější, a to minimalizací součtu čtvercových rozdílů v rámci každého shluku.



Obrázek 42 Hierarchické shlukování v programu Orange.

Aplikací tohoto nastavení vzniklo celkem **12 shluků**. 27 měst se nachází v celkem 12 skupinách po dvou nebo třech městech. Ostatních 77 měst se při tomto nastavení nenachází v žádném shluku a vystupují samostatně (vlivem nastavené prahové vzdálenosti). Celý dendrogramu je součástí příloh (příloha: Vystupni_Data/IMG/Europe100_HierarchicalClustering.png). Při nastavení kratší vzdálenosti není každé město zařazeno do shluku, ale zároveň jsou vzniklé shluky homogennější a výstižnější. Vzniklé shluky měst (Obrázek 43) byly dále zkoumány.



Obrázek 43 Hierarchické shlukování – graf siluety jednotlivých shluků.

Intepretace shluků

Pro každý se shluků byly vybrány nejtypičtější frekventované sady, a to pomocí manuálního procházení **krabicových grafů (box plotů) v programu Orange**, které zobrazovaly vždy pro vybranou frekventovanou sadu průměrnou podporu v rámci shluku a průměrnou hodnotu v rámci ostatních měst společně se Studentových t-testem a hodnotou p pro usuzování statistické významnosti (Obrázek 44).

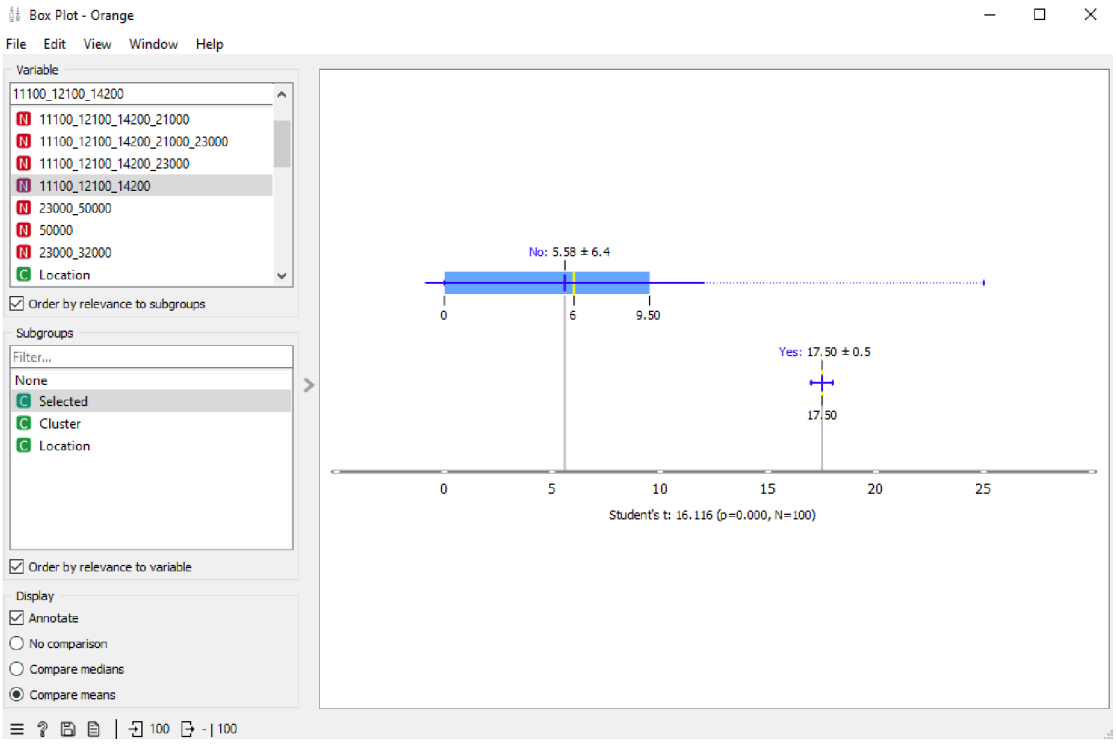
Funkce "**Uspořádat podle významu pro podskupiny**" v programu Orange je nástroj, který je k dispozici v různých widgetech pro analýzu dat. Ve widgetu krabicový graf (box plot) umožňuje seřadit data na základě toho, jak dobře mohou hodnoty v určitém sloupci rozlišovat mezi podskupinami přítomnými v datech.

Tato funkce funguje tak, že pro každý sloupec vypočítá skóre relevance, které měří sílu asociace mezi hodnotami sloupce a podskupinami v datech. Skóre relevance se vypočítá pomocí statistického testu, který je vhodný pro typ dat ve sloupci. Konkrétně pokud sloupec obsahuje kategoriální údaje, jako jsou nominální nebo ordinální proměnné, použije se k výpočtu skóre relevance test chí-kvadrát. Pokud sloupec obsahuje spojité údaje, například číselné proměnné, použije se test ANOVA.

Test chí-kvadrát porovnává pozorované a očekávané četnosti jednotlivých kategorií ve sloupci, aby se posoudilo, zda existují významné rozdíly v těchto četnostech mezi podskupinami. Test **ANOVA** naproti tomu porovnává průměry jednotlivých podskupin ve sloupci, aby zjistil, zda se tyto průměry od sebe významně liší.

Po výpočtu skóre významnosti pro každý sloupec seřadí Orange data sestupně na základě těchto skóre. To umožňuje uživateli určit sloupce, které mají nejvyšší relevanci pro podskupiny, a jsou tedy nejúčinnější při rozlišování mezi podskupinami v datech.

Tato metoda umožňuje nalézt sloupce (frekventované sady), kterými se daný shluk měst odlišuje od zbývajících měst. Jinými slovy, které frekventované sady jsou pro daný shluk typické a nejvíce zapříčinily zařazení daných měst do odpovídajících shluků.



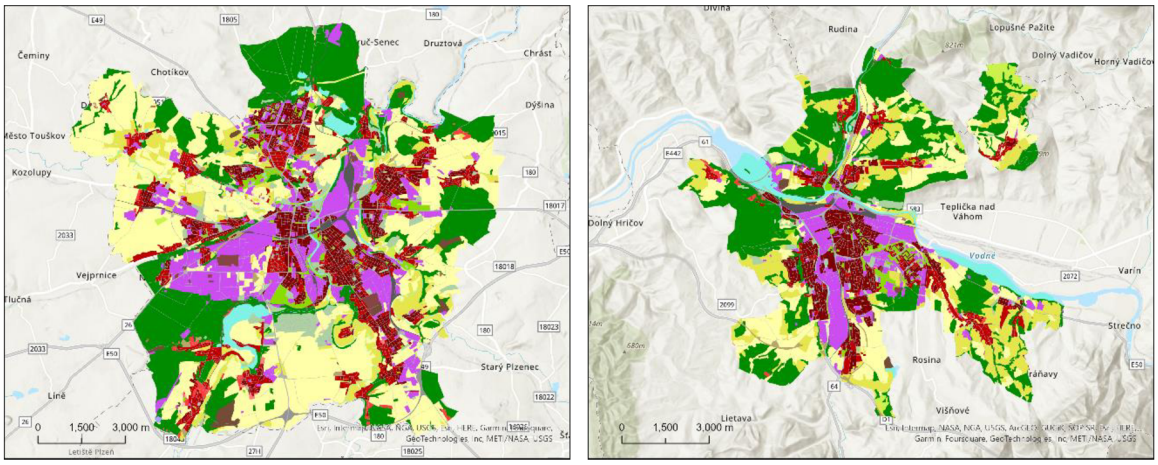
Obrázek 44 Krabicový graf frekventované sady 11100_12100_14200 pro shluk měst Plzeň a Žilina.

Dále je potřeba brát při interpretaci skupin v úvahu, že se nejedná o obrazovou podobnost. Obdobná podpora určité frekventované sady sousednosti nevypovídá nic o prostorovém rozložení jednotlivých využití území. Například, když je nalezena frekventovaná sada s vysokou podporou pro dvě města, může to znamenat pouze to, že určité využití území **se vyskytuje v těchto oblastech s podobnou frekvencí**. Nicméně prostorové uspořádání, velikost a další charakteristiky mohou být naprosto odlišné. Frekvence ale může být v závislosti na velikosti města v absolutních hodnotách jiná.

Je tedy důležité, aby se výsledky analýzy frekventovaných sad interpretovaly opatrně a zohledňovaly další faktory. Analytické metody založené na lokálních vzorech využití území mohou **poskytnout užitečné informace o trendech a vzorech využití území**, ale je důležité si být vědom omezení těchto metod a interpretovat výsledky s ohledem na konkrétní kontext. V rámci jednotlivých shluků byly identifikovány potencionální typické frekventované sady, které byly statisticky porovnány s ostatními městy. Součástí popisu je výčet typických frekventovaných sad a jejich procentuální podpory.

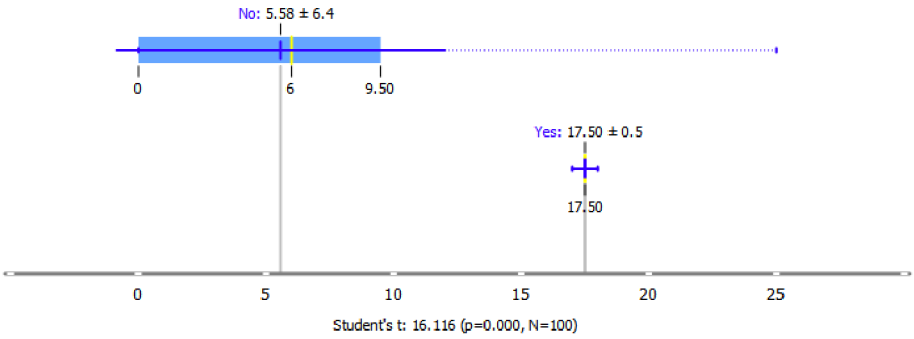
5.3.1 Plzeň (CZ), Žilina (SK) – C1

Plzeň je město ležící v západních Čechách, v blízkosti řeky Radbuzy a Berounky. Je to čtvrté největší město v České republice s počtem obyvatel přes 170 000. Plzeň je historicky významná jako centrum pivovarnictví. Město je také známé svými historickými památkami. Plzeň je důležitým průmyslovým a obchodním centrem regionu. **Žilina** je město na severním Slovensku, ležící v údolí řeky Váhu. Město má přibližně 80 000 obyvatel. Je to důležité průmyslové město s dlouhou historií a mnoha kulturními památkami. Žilina je také důležitým dopravním uzlem, s dálnicí D1 a železniční tratí spojující město s ostatními částmi Slovenska a Polska. Město je také výchozím bodem pro výlety do slovenských hor.



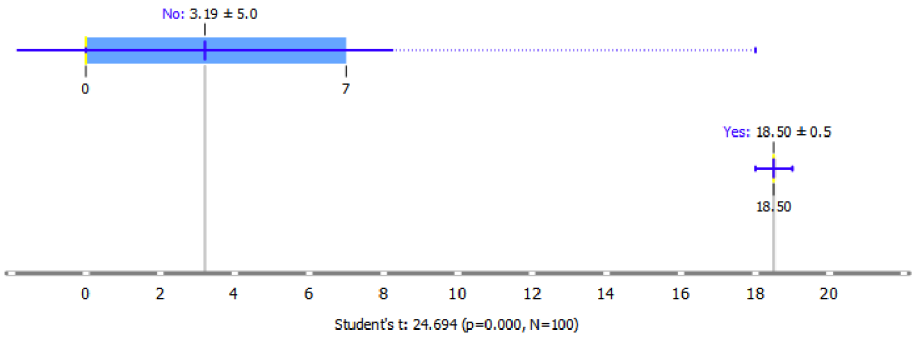
Obrázek 45 Náhled – vlevo Plzeň (CZ) a vpravo Žilina (SK) (zdroj: data UA 2018).

Centra obou měst se vyznačují **souvislou hustou městskou zástavbou (11100), průmyslovými, komerčními a veřejnými plochami (12100) a sportovišti (14200)**. 11100 12100 14200 Ve městě Plzeň má tato sada podporu 18 % a v Žilině 17 %. Průměrná hodnota podpory v rámci shluku je 17,5 %. U ostatních měst v rámci sady 100 evropských měst je podpora tohoto pravidla v průměru 5,58 % ± 6,4 (Obrázek 46). Hodnota Studentova t-testu 16,116 ukazuje, že mezi průměry obou srovnávaných skupin je velký rozdíl, který je statisticky významný. Absolutní hodnota t-hodnoty je velká, což naznačuje, že rozdíl mezi průměry pravděpodobně není způsoben pouze náhodou. Hodnota p rovna 0,000 naznačuje, že rozdíl mezi průměry je vysoce statisticky významný (pozn. Orange zaokrouhluje hodnotu p na 3 desetinná místa, hodnota 0,000 neznamená, že se hodnota p rovná přesně hodnotě nula. Interpretace takového hodnoty je, že se jedná o hodnotu velmi malou, ne však nulovou).



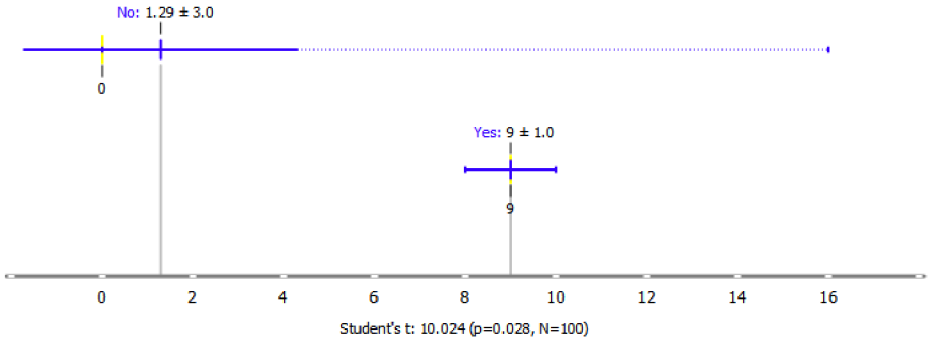
Obrázek 46 FS souvislé husté městské zástavby (11100), průmyslových, komerčních a veřejných ploch (12100) a sportovišť (14200).

Přechod z urbánního centra města do rurálního okolí se vyznačuje **nesouvislou hustou městskou zástavbou (11210)** přecházející v **ornou půdou (21000) a lesy (31000)** nacházející se v sousedství. 11210 21000 31000 Ve městě Plzeň má tato sada podporu 19 % a v Žilině 18 %. Průměrná hodnota podpory v rámci shluku je 18,5 %. U ostatních měst v rámci sady 100 evropských měst je podpora tohoto pravidla v průměru 3,19 % ± 5,0 (Obrázek 47). Hodnota t-testu je rovna 24,694 s hodnotou p 0,000.



Obrázek 47 FS nesouvislé husté městské zástavby (11210), orné půdy (21000) a lesa (31000).

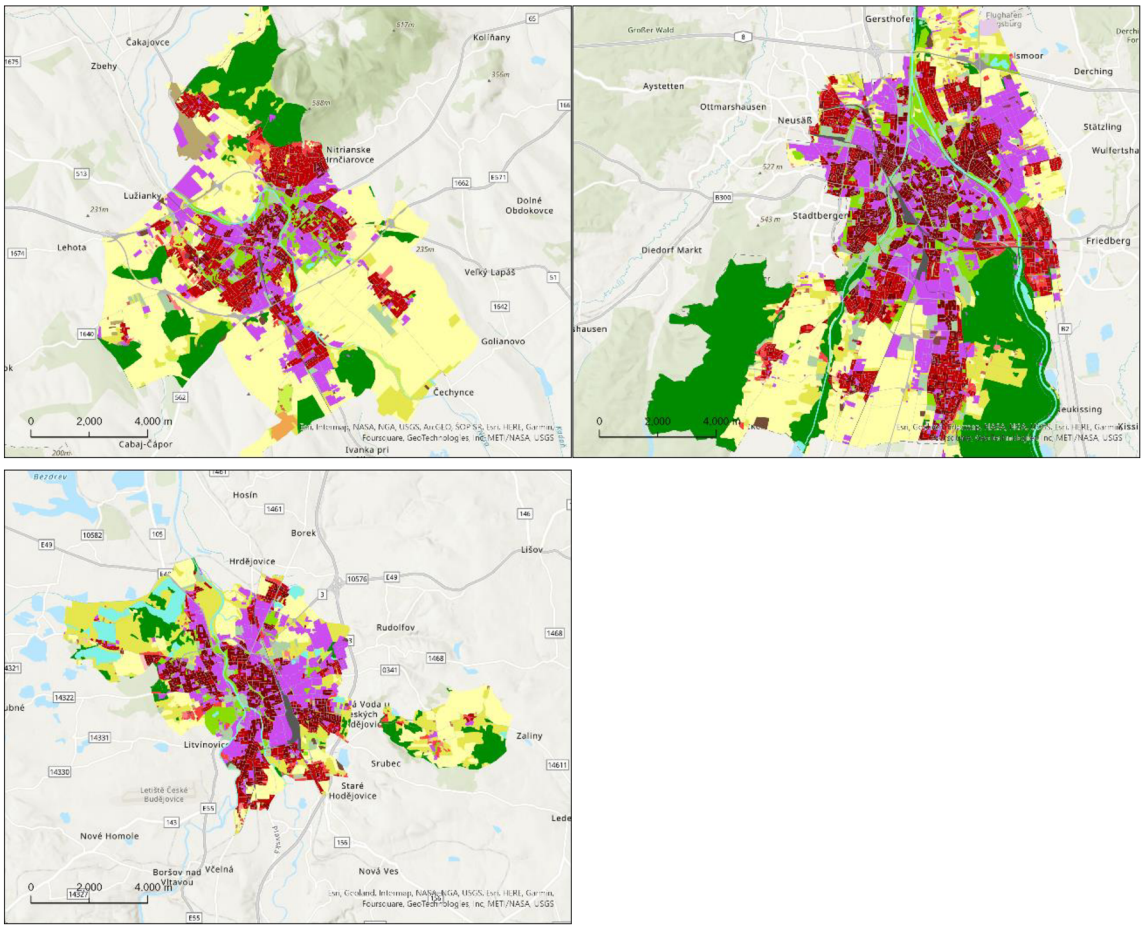
Na území obou měst se typicky vyskytuje **vodní plocha (50000)** sousedící s **ornou půdou (21000)** a **lesy (31000)** 21000 31000 50000 s průměrnou podporou 9 % v rámci shluku. U ostatních měst se tato sada vyskytuje s průměrnou podporou 1,29 % ± 3,0 (Obrázek 48).



Obrázek 48 FS orné půdy (21000), lesů (31000) a vodní plochy (50000).

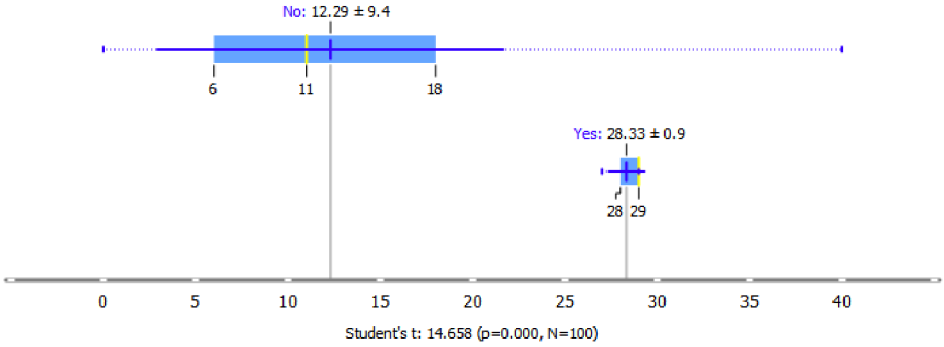
5.3.2 Nitra (SK), České Budějovice (CZ), Augsburg (DE) – C2

Nitra je město v západním Slovensku, ležící na břehu řeky Nitra. Město má přibližně 80 000 obyvatel a je to významné kulturní a hospodářské centrum regionu. Mezi nejvýznamnější památky patří Nitranský hrad a historické centrum města. **České Budějovice** jsou město v jižních Čechách, ležící na břehu řeky Vltavy. Město má přibližně 100 000 obyvatel a je to významné hospodářské a kulturní centrum regionu. České Budějovice jsou také známé svým pivem, Budějovickým Budvarem, které je exportováno do celého světa. **Augsburg** je město v jižním Německu, ležící na řece Lech. Město má přibližně 300 000 obyvatel a je to významné kulturní a hospodářské centrum regionu. Augsburg je známý svou dlouhou historií a mnoha památkami. Město je také důležitým průmyslovým centrem, s významnými firmami jako jsou Siemens a MAN.



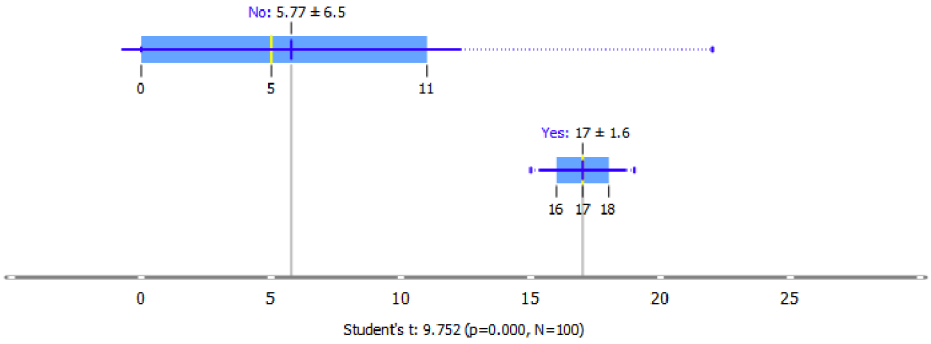
Obrázek 49 Náhled – vlevo nahoře Nitra (SK), vpravo nahoře Augsburg, vlevo dole České Budějovice (zdroj: UA 2018).

Centra měst se vyznačují mozaikou vyznačující se **sousedností souvislé husté městské zástavby (11100)**, **průmyslových, komerčních a veřejných ploch (12100) a městské zeleně (14100)**. 12100 11100 14100 Ve městě Nitra má tato sada podporu 29 %, v Augsburgu má podporu 29 % a v Českých Budějovicích 27 %. Průměrná hodnota podpory v rámci shluku je 28,33 % ± 0,9. U ostatních měst v rámci sady 100 evropských měst je podpora tohoto pravidla v průměru 12,29 % ± 9,4 (Obrázek 50). Hodnota Studentova t-testu 14,658 ukazuje, že mezi průměry obou srovnávaných skupin je značný rozdíl, který je statisticky významný. Hodnota p rovna 0,000 naznačuje, že rozdíl mezi průměry je vysoce statisticky významný.



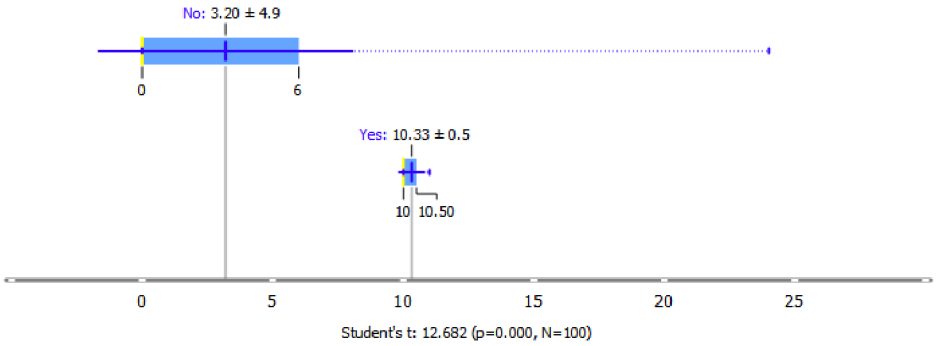
Obrázek 50 FS souvislé husté městské zástavby (11100), průmyslové, komerční a veřejné plochy (12100) a veřejné zeleně (14100).

V okolí centra města se nachází v sousedství **souvislá hustá městská zástavba (11100) s nesouvislou hustou městskou zástavbou (11210) a sportovišti (14200)**. 11100 11210 14200 Průměrná podpora pravidla v rámci shluku je 17 % ± 1,6 (Obrázek 51).



Obrázek 51 FS souvislé husté městské zástavby (11100), nesouvislé husté městské zástavby (11210) a sportoviště (14200).

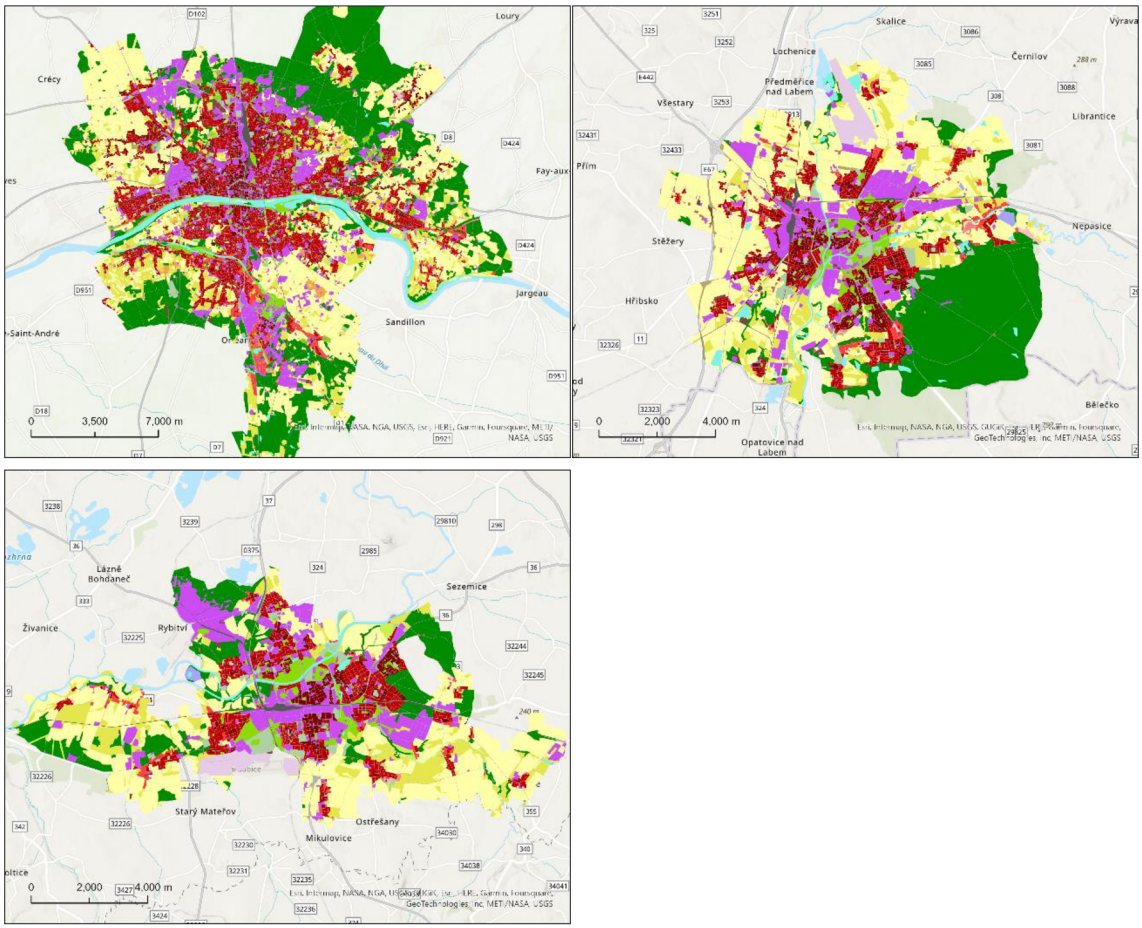
Poměrně časté je sousedství 14100 50000 **městské zeleně (14100) s vodní plochou (50000)**, které se vyskytuje v rámci shluku s průměrnou podporou 10,33 % ± 0,5 (Obrázek 52). Centrem měst prochází vodní toky v jejichž blízkosti je městská zeleň.



Obrázek 52 FS městské zeleně (14100) a vodní plochy (50000).

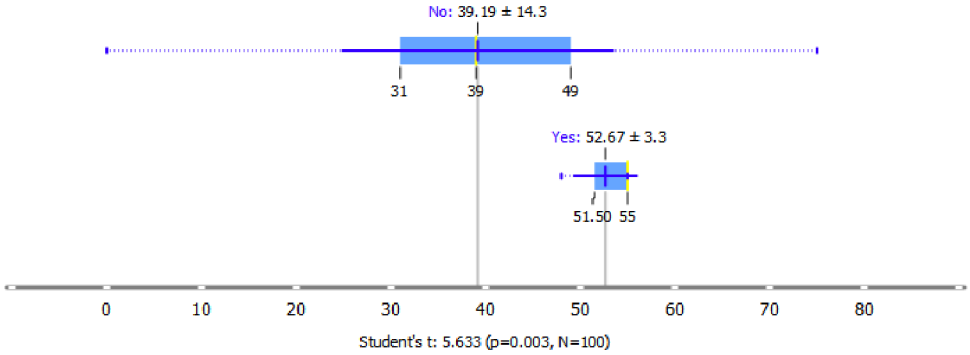
5.3.3 Orleans (FR), Hradec Králové (CZ), Pardubice (CZ) – C3

Orléans je město ve střední Francii, ležící na řece Loiře. Město je významné historické centrum. Počet obyvatel je přibližně 115 000. Mezi nejvýznamnější památky v Orléansu patří katedrála svatého Kříže, gotický kostel sv. Pavla a historické centrum města, které je plné malebných uliček a náměstí. **Hradec Králové** je město východních Čech, ležící na soutoku řek Labe a Orlice. Město má přibližně 92 000 obyvatel a je to významné kulturní a hospodářské centrum regionu. Hradec Králové je také významným turistickým cílem, díky svým historickým památkám. **Pardubice** jsou město ve východních Čechách, ležící na řece Labe. Město má přibližně 93 000 obyvatel a je to významné hospodářské a kulturní centrum regionu. Město je významným železničním uzlem, s důležitými tratěmi vedoucími do Prahy a dalších částí České republiky.



Obrázek 53 Náhled – vlevo nahoře Orleans (FR), vpravo nahoře Hradec Králové (CZ), vlevo dole Pardubice (CZ) (zdroj: UA 2018).

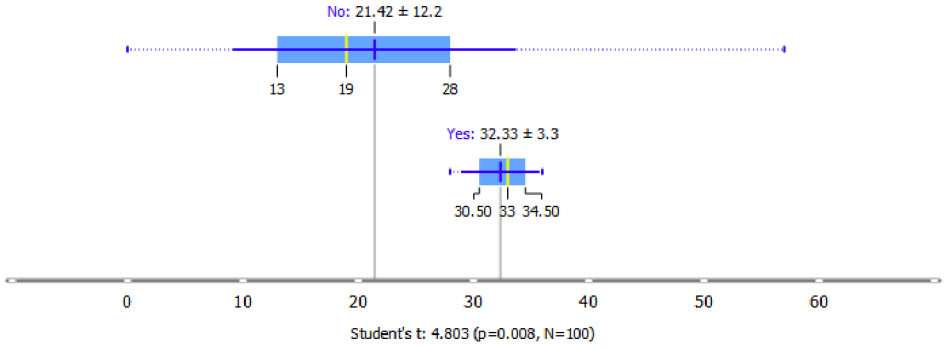
V porovnání s předchozím shlukem C2, ve kterém dominuje souvislá hustá městská zástavba (11100) nad **nesouvislou hustou zástavbou (11210)** v tomto shluku je to právě naopak. Celkově se ale jedná o poměrně podobné skupiny měst. Významně zastoupenou sousedností je sousednost již zmíněné **nesouvislé husté městské zástavby s průmyslovými, komerčními a veřejnými plochami (12100)** 12100 11210 s průměrnou podporou 52,67 % v rámci shluku. U ostatních měst je podpora nižší, v průměru 39,19 % ± 14,3. Rozptyl podpory napříč městy je poměrně velký, hodnota t-testu je nižší než u předchozích frekventovaných sad.



Obrázek 54 FS nesouvislé husté městské zástavby (11210) a průmyslových, komerčních a veřejných ploch (12100).

Další frekventovaná sada podporuje myšlenku, že v rámci shluku je typičtější **nesouvislá hustá zástavba (11210)** na úkor souvislé husté městské zástavby (11100). Tento typ zástavby se vyskytuje také v kombinaci **s městskou zelení (14100)** s průměrnou podporou 32,33 % ± 3,3. Průměrná podpora v rámci datasetu je 21,42 % ± 12,2 (Obrázek 55).

11210 14100

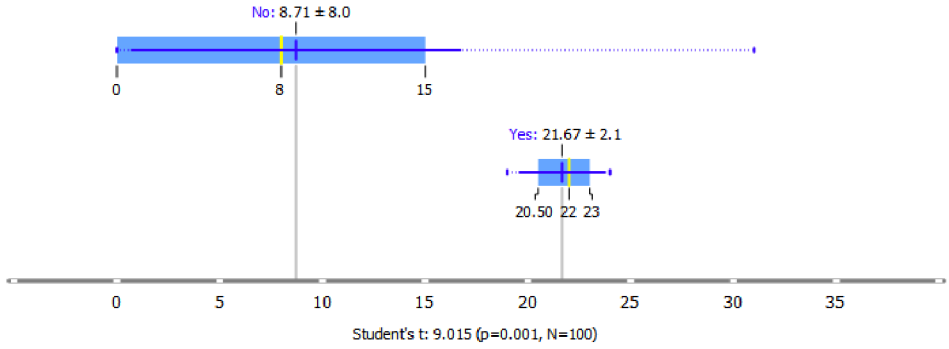


Obrázek 55 FS nesouvislé husté městské zástavby (11210) s městskou zelení (14100).

Přechod z urbánního centra města do rurálního okolí se vyznačuje **nesouvislou hustou městskou zástavbou (11210)** přecházející v **ornou půdou (21000) a pastviny (31000) nacházející se v sousedství**.

11210 21000 23000

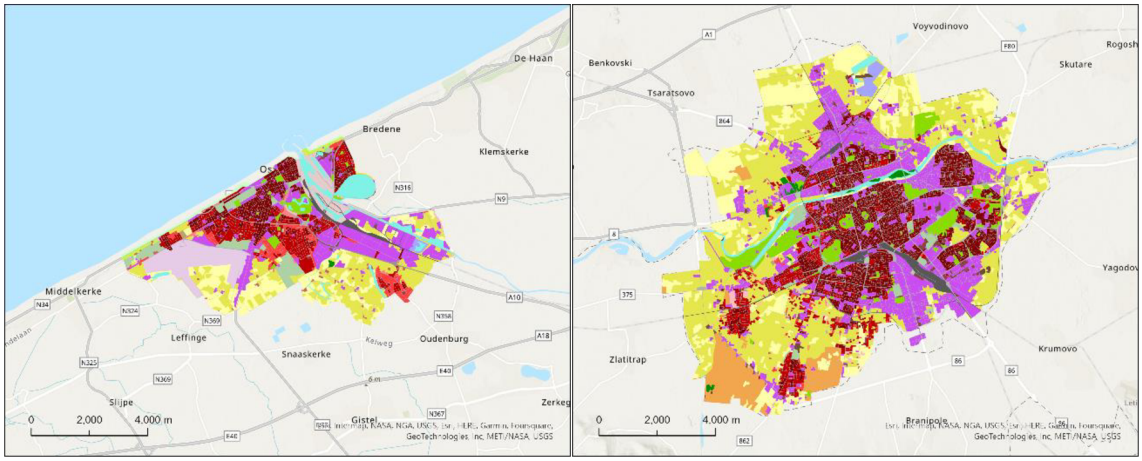
Průměrná podpora této sady je 21,67 % ± 2,1. U ostatních měst v rámci datasetu je podpora tohoto pravidla v průměru 8,71 % ± 8,0 (Obrázek 56). Hodnota t-testu je rovna 9,015 s hodnotou p 0,001.



Obrázek 56 FS nesouvislé husté městské zástavby (11210), orné půdy (21000) a pastvin (23000).

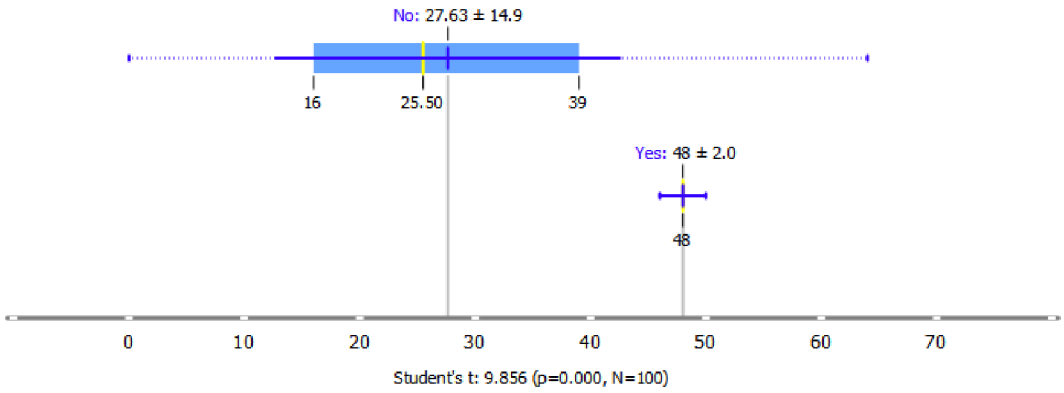
5.3.4 Oostende (BE), Plovdiv (BG) – C4

Plovdiv má přibližně 340 000 obyvatel a je to druhé největší město v Bulharsku po Sofii. Leží v jižní části země na obou březích řeky Maricy. Město má bohatou historii sahající až do starověku, což je zřejmé z mnoha dochovaných památek. **Oostende** je město v Belgii, ležící na pobřeží Severního moře. Město má přibližně 70 000 obyvatel a je populárním turistickým letoviskem, díky svým plážím a promenádě. Oostende má také bohatou historii a kulturu, která se odráží v mnoha památkách a muzeích.



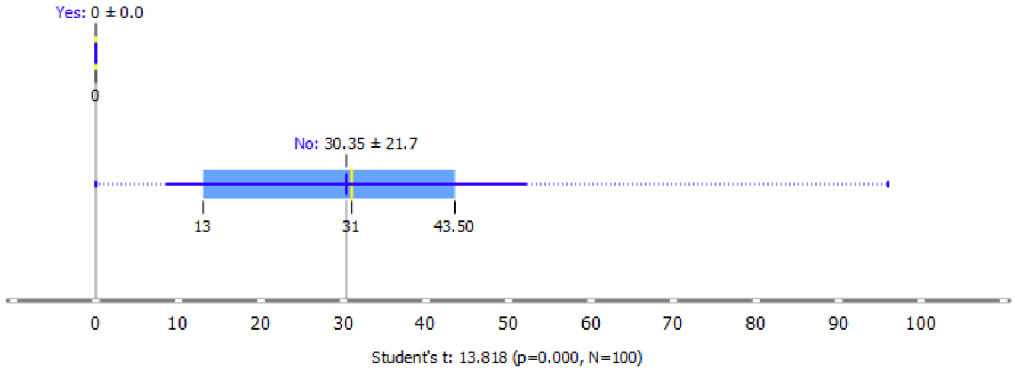
Obrázek 57 Náhled – vlevo Oostende (BE), vpravo Plovdiv (BG) (zdroj: UA 2018).

Města v rámci shluku mají významně odlišný počet obyvatel, Plovdiv má přibližně pětkrát tolik obyvatel než Oostende. Rozdílný je tím pádem i počet polygonů využití území, který bude u Plovdivu větší. První ze dvou zjištěných frekventovaných sad sousednosti, kterou se shlukovaná města odlišují od ostatních je kombinace **souvislé husté městské zástavby (11100) s průmyslovými, komerčními a veřejnými plochami (12100)**. 12100 11100 Průměrná hodnota podpory pravidla je 48 % ± 2,0. Hodnota Studentova t-testu je 9,856 s p-hodnotou 0,000, tedy se jedná o statisticky významnou frekventovanou sadu v rámci shluku.



Obrázek 58 FS souvislé husté městské zástavby (11100) s průmyslovými, komerčními a veřejnými plochami (12100).

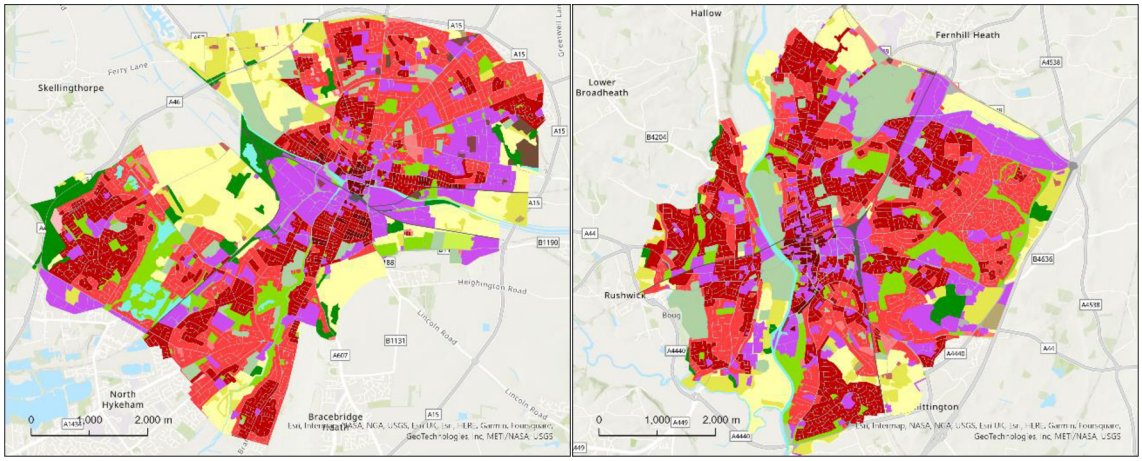
Při interpretaci je potřeba brát v potaz také negativní pravidla. Při bližším pohledu na shlukovaná města je patrná absence **lesů (31000)**, 31000 které v rámci urbánního jádra Plovdivu a Oostende při minimální podpoře 5 % netvoří žádné frekventované sady sousedství. U ostatních měst je podpora pravidel obsahujících lesy značně nekonzistentní s průměrnou hodnotou 30,35 %± 21,7.



Obrázek 59 FS lesy (31000).

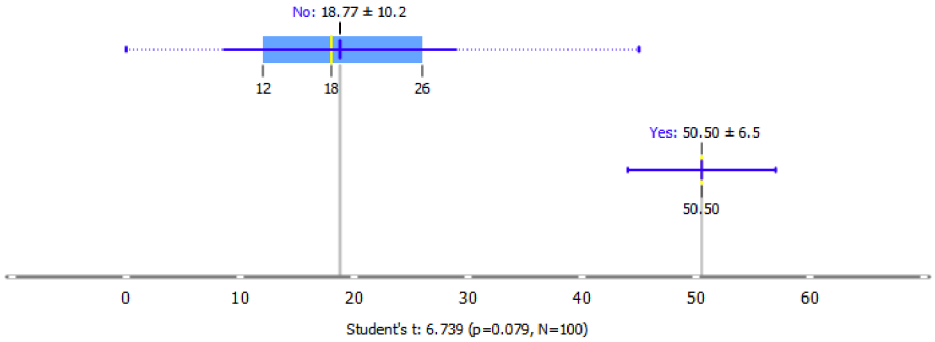
5.3.5 Lincoln (UK), Worcester (UK) – C5

Další skupinou podobných měst nalezených pomocí hierarchického shlukování je dvojice britských měst Lincoln a Worcester. **Lincoln** je historické město v centrální Anglii s populací přibližně 100 000 obyvatel. Je to významné kulturní a turistické centrum s množstvím historických památek. **Worcester** je historické město v Anglii, ležící na řece Severn v hrabství Worcestershire. Město má přibližně 100 000 obyvatel a je známé svou bohatou historií a kulturou.



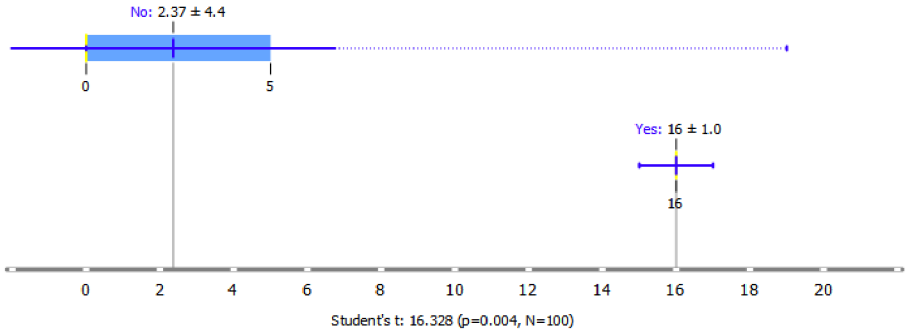
Obrázek 60 Náhled – vlevo Lincoln (UK), vpravo Worcester (UK) (zdroj: UA 2018).

Vyznačují se typickou britskou **nesouvislou středně hustou zástavbou (11220) v kombinaci s nesouvislou hustou městskou zástavbou (11210) a průmyslovými, komerčními a veřejnými plochami (12100)** . 11210 12100 11220 Ve městě Worcester má tato sada podporu 57 % a v Lincolnu 44 %. Průměrná hodnota podpory v rámci shluku je 50,5 %. U ostatních měst v rámci sady 100 evropských měst je podpora tohoto pravidla v průměru 18,77 (Obrázek 61). Hodnota Studentova t-testu 6,739 ukazuje, že mezi průměry obou srovnávaných skupin je rozdíl, který je statisticky významný.



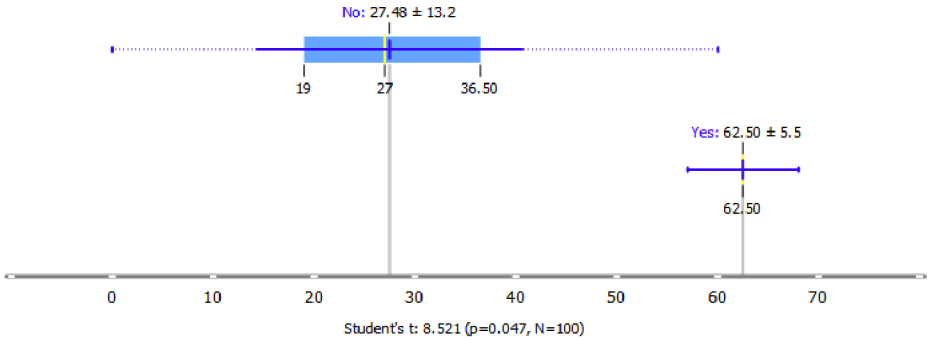
Obrázek 61 FS nesouvislé husté městské zástavby (11210), nesouvislé středně husté zástavby (11220) a průmyslových, komerčních a veřejných ploch (12100).

Významně se tato dvojice měst odlišuje od ostatních přítomností frekventované sady sousedství **nesouvislé středně husté zástavby (11220) s městskou zelení (14100) a sportovišti (14200)**. 11220 14100 14200 Průměrná podpora této sady je 16 % ± 1,0. Průměrná podpora napříč ostatními městy je 2,37 % ± 4,4. Studentův t-test s hodnotou 16,328 a hodnota p 0,004 indikuje statistickou významnost této sady (Obrázek 62).



Obrázek 62 FS nesouvislé středně husté městské zástavby (11220) s městskou zelení (14100) a sportovišti (14200)

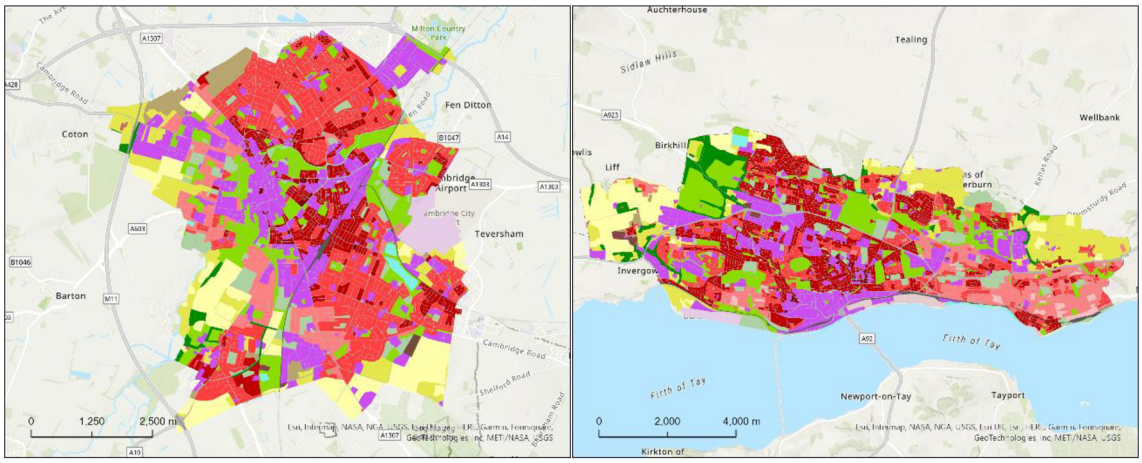
Razantně zastoupenější je frekventovaná sada sousednosti nesouvislé husté městské zástavby (11210) a nesouvislé středně husté městské zástavby (11220). 11210 11220 Průměrná podpora pravidla v rámci datasetu je 62,5 % ± 5,5. Průměrná podpora sady u ostatních měst je 27,48 % ± 13,2.



Obrázek 63 FS nesouvislé husté městské zástavby (11210) a nesouvislé středně husté městské zástavby (11220).

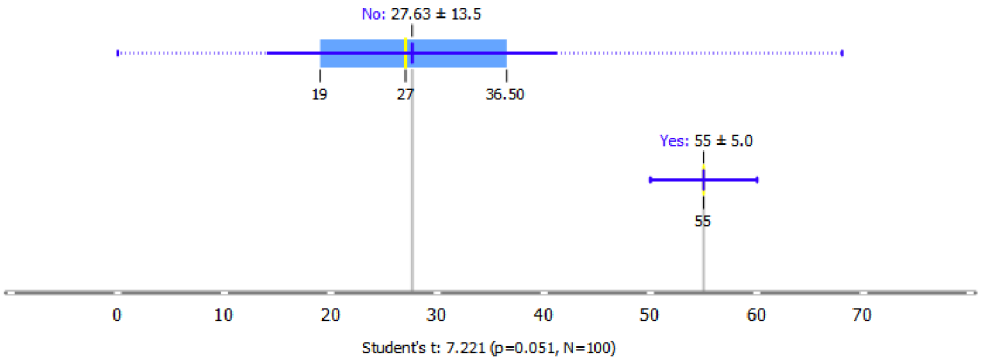
5.3.6 Cambridge (UK), Dundee City (UK) – C6

Cambridge je univerzitní město v Anglii, ležící v hrabství Cambridgeshire na řece Cam. Město má přibližně 140 000 obyvatel a je známé díky svému významu v oblasti vzdělání a výzkumu. Univerzita v Cambridge je jednou z nejstarších a nejuznávanějších v Evropě. Město je plné historických budov, jako je například katedrála svatého Jana Evangelisty a univerzitní knihovna. Cambridge je také oblíbeným turistickým cílem, především díky svým krásným parkům a zahradám. **Dundee City** je město v severovýchodním pobřežním regionu Skotska. Město má přibližně 150 000 obyvatel a je čtvrtým největším městem ve Skotsku. Dundee má bohatou historii, která sahá až do 12. století, a v minulosti bylo důležitým přístavem a centrem textilního průmyslu. Dnes je Dundee významným centrem výzkumu a vzdělání, s univerzitou Dundee. Dundee se také nachází na úpatí kopce Law.



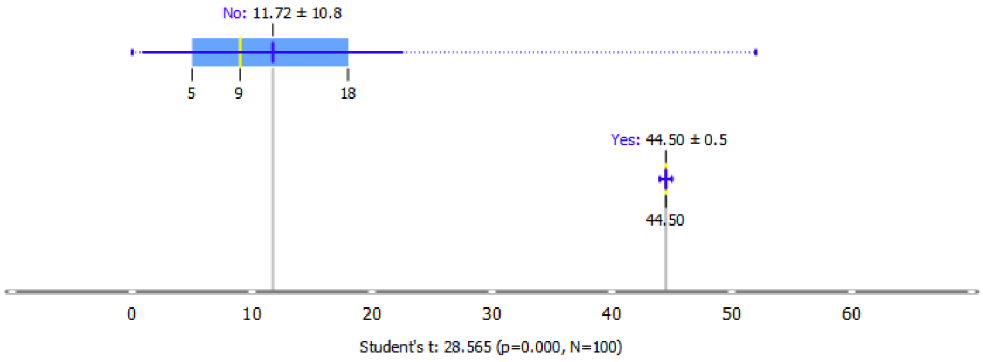
Obrázek 64 Náhled – vlevo Cambridge (UK), vpravo Dundee City (UK) (zdroj: UA 2018).

Tato města se opět vyznačují typickou britskou **nesouvislou středně hustou zástavbou (11220)** v kombinaci s **nesouvislou hustou městskou zástavbou (11210)**. 11220 11210 Ve městě Cambridge má tato sada podporu 50 % a v Dundee City 60 %. Průměrná hodnota podpory v rámci shluku je 55 %. U ostatních měst v rámci sady 100 evropských měst je podpora tohoto pravidla v průměru 27,63 % (Obrázek 65). Hodnota Studentova t-testu 7,221 ukazuje, že mezi průměry obou srovnávaných skupin je rozdíl, který je statisticky významný. Hodnota p je 0,051.



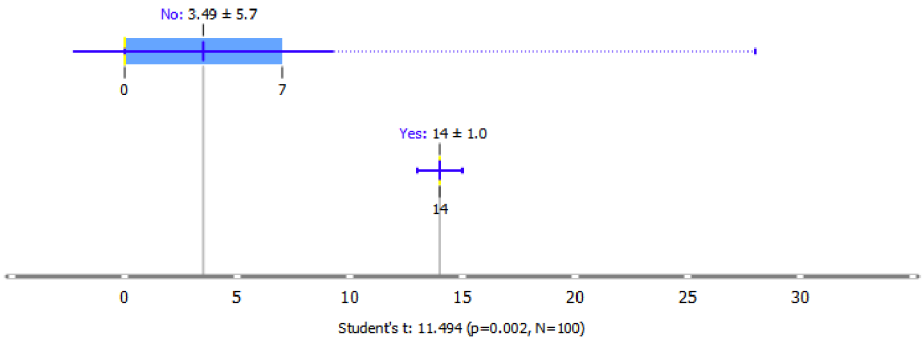
Obrázek 65 FS nesouvislé husté městské zástavby (11210) a nesouvislé středně husté městské zástavby (11220).

Další frekventovanou sadou, kterou se města ve shluku odlišují je sousedství **nesouvislé středně husté zástavby (11220) a městské zeleně (14100).** 11220 14100 Průměrná podpora pravidla je 44,5 % ± 0,5. U ostatních měst je podpora významně nižší s průměrem 11,72 % ± 10,8. Hodnota Studentova t-testu nabývá hodnoty 28,565 a hodnota p 0,000 což poukazuje na statisticky významnou odlišnost průměrných hodnot (Obrázek 66).



Obrázek 66 FS nesouvislé středně husté zástavby (11220) a městské zeleně (14100).

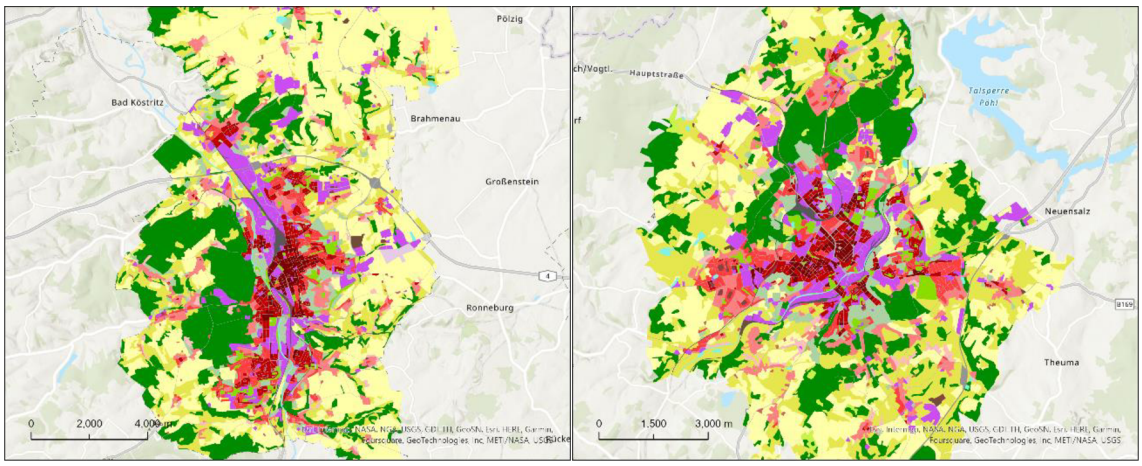
Druhou obdobnou frekventovanou sadou sousednosti je kombinace **nesouvislé málo husté městské zástavby (11230) se sportovišti (14200).** 14200 11230 Průměrná podpora pravidla nabývá hodnoty 14 % ± 1,0. Zbývající města mají podporu sady nižší v průměru 3,49 % ± 5,7 (Obrázek 67).



Obrázek 67 FS nesouvislé málo husté městské zástavby (11230) se sportovišti (14200).

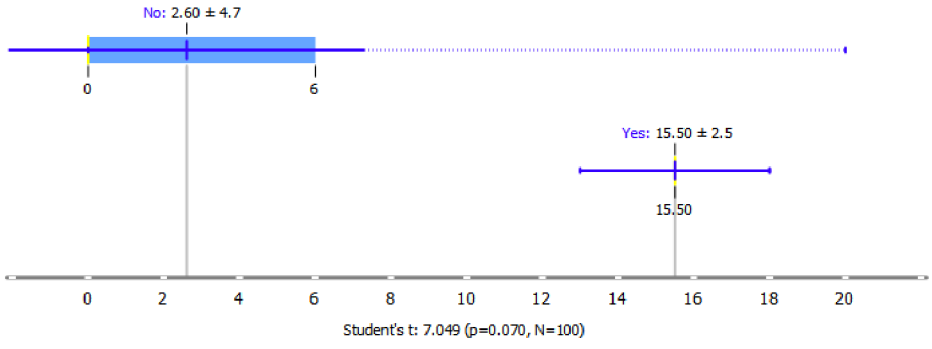
5.3.7 Plauen (DE), Gera (DE) – C7

Plauen a Gera jsou města v Německu, nacházející se v regionu Sasko. **Plauen** je městem s přibližně 65 000 obyvateli a je známé díky své textilní historii a průmyslu. V minulosti bylo Plauen jedním z nejvýznamnějších center textilní výroby v Evropě. Dnes je město turistickým cílem díky svým historickým budovám. Plauen se nachází v malebném krajině s mnoha lesy a jezer, což nabízí příležitosti pro turistiku a rekreaci. **Gera** je město s přibližně 95 000 obyvateli a je známé díky své historii v oblasti průmyslu a výroby. Mezi nejvýznamnější památky patří například hrad Osterstein. Gera je také důležitým kulturním centrem s mnoha divadly a muzei. Město se nachází v údolí řeky Weiße Elster a okolní krajinu tvoří kopce a lesy, což nabízí příležitosti pro turistiku a outdoorové aktivity.



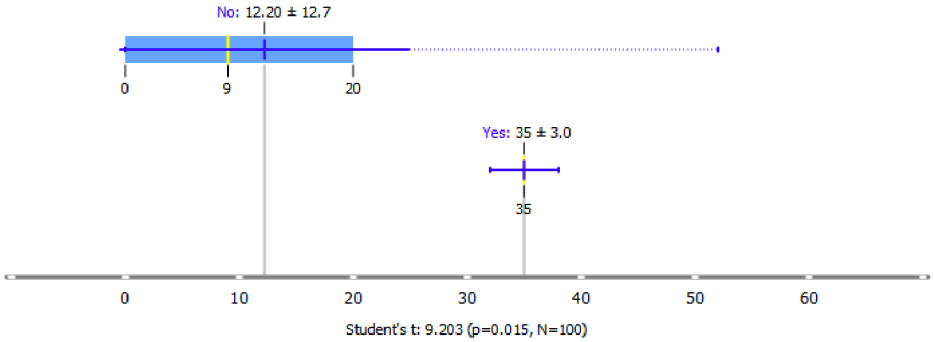
Obrázek 68 Náhled – vlevo Plauen (DE), vpravo Gera (DE) (zdroj: UA 2018).

V rámci analýzy sousednosti měst Gera a Plauen byly nalezeny 4 významné frekventované sady. První z nich je sousedství **nesouvislé málo husté městské zástavby (11230), nesouvislé velmi málo husté městské zástavby (11240), orné půdy (21000) a pastviny (23000).** 23000 21000 11230 11240 Tento typ sousedství se vyskytuje při přechodu z urbánního centra města, které se vyznačuje hustší zástavbou, do rurálního okolí. Průměrná podpora této frekventované sady činí 15,5 % ± 2,5. U ostatních měst je podpora této kombinace nižší (průměr 2,6 % ± 4,7) (Obrázek 69).



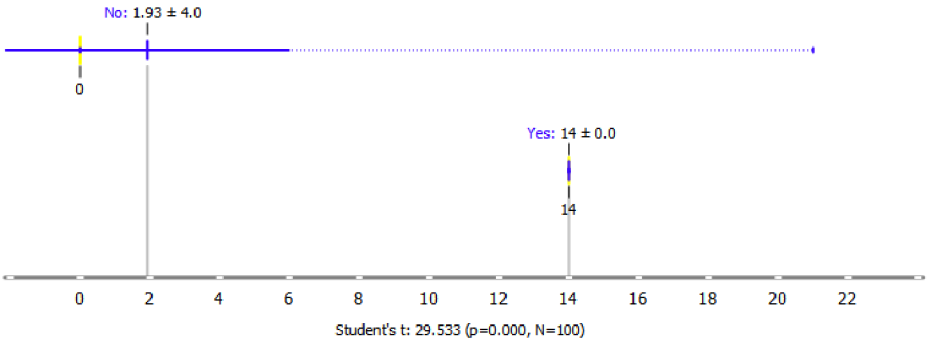
Obrázek 69 FS nesouvislé málo husté městské zástavby (11230), nesouvislé velmi málo husté městské zástavby (11240), orné půdy (21000) a pastvin (23000).

Okolí města je typické kombinací **orné půdy (21000), pastvin (23000) a lesů (31000)**, které se navzájem mísí a tvoří typickou trojbarevnou mozaiku. 23000 21000 31000 Podpora tohoto typu sousedství nabývá v rámci shluku průměrné hodnoty 35 % ± 3,0.



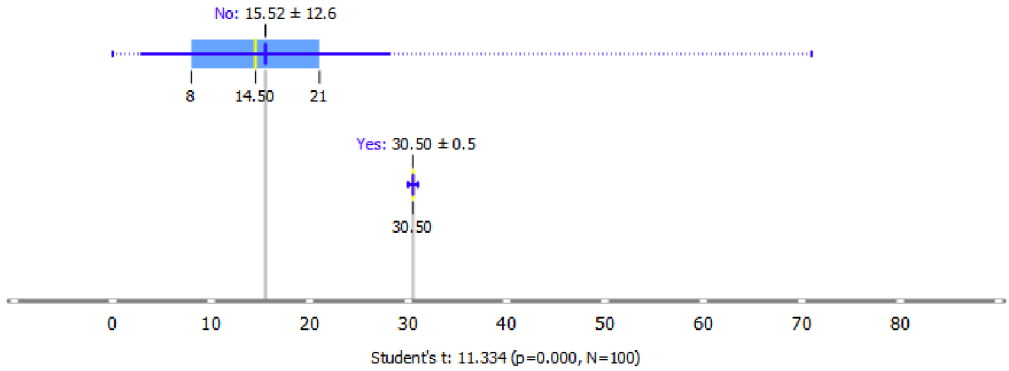
Obrázek 70 FS orné půdy (21000), pastvin (23000) a lesů (31000).

Centrum města je typické dvěma kombinacemi sousedství využití území. První z nich obsahuje **nesouvislou málo hustou městskou zástavbu (11230) v kombinaci s průmyslovými, komerčními a veřejnými plochami (12100) a sportovišti (14200)**. 12100 11230 14200 Průměrná podpora sady je 14 %. U ostatních měst je podpora nižší a to 1,93 % ± 4,0. Studentův t-test i hodnota p ukazují na významnou statistickou významnost rozdílů průměrů (Obrázek 71).



Obrázek 71 FS nesouvislé málo husté městské zástavby (11230), průmyslových, komerčních a veřejných ploch (12100) a sportovišť (14200).

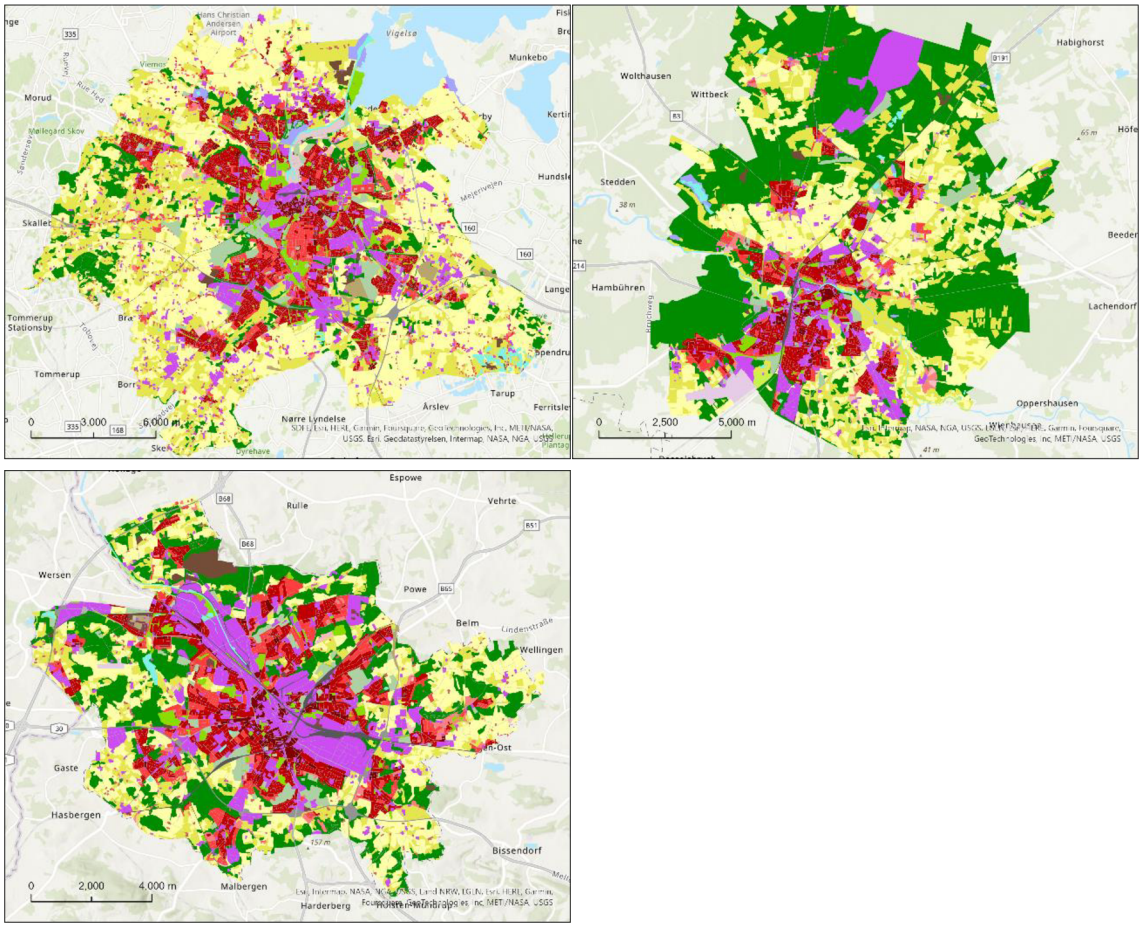
Druhou z nich je sousedství **nesouvislé středně husté městské zástavby (11220) a nesouvislé málo husté městské zástavby (11230).** 11230 11220 Tato sada má v průměru podporu 30,5 % ± 0,5. Studentův t-test i hodnota p ukazují na významnou statistickou významnost rozdílů průměrů (Obrázek 72).



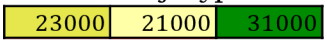
Obrázek 72 FS nesouvislé středně husté městské zástavby (11220) a nesouvislé málo husté městské zástavby (11230).

5.3.8 Odense (DK), Celle (DE) Osnabruck (DE) – C8

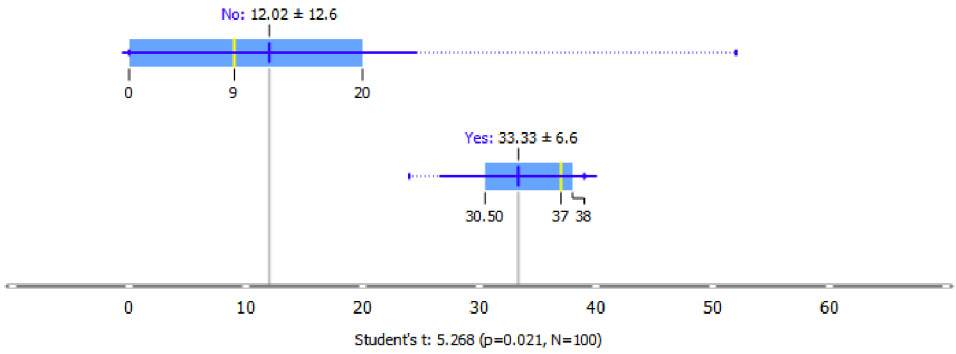
Odense je třetím největším městem Dánska s počtem obyvatel přibližně 200 000. Je to rodiště známého spisovatele Hanse Christiana Andersena a město je také známé pro svůj středověký střed s gotickou katedrálou a hradem. Město má také mnoho muzeí, galerií a parků a je důležitým kulturním centrem Dánska. **Celle** je menší německé město s počtem obyvatel přibližně 70 000, nacházející se v Dolním Sasku. Město je známé pro svůj historický střed s mnoha malebnými uličkami a hrázděnými domy. Celle se nachází v krajině s mnoha lesy a jezery, což nabízí příležitosti pro turistiku a rekreaci. **Osnabrück** je město v Dolním Sasku s počtem obyvatel přibližně 170 000 a je jedním z nejdůležitějších měst v regionu. Je známé pro svůj historický střed s mnoha gotickými budovami a katedrálou svatého Petra. Osnabrück se nachází v blízkosti pohoří Teutoburský les a má mnoho parků a zelených ploch, což nabízí příležitosti pro turistiku a rekreaci v přírodě.



Obrázek 73 Náhled – vlevo nahoře Odense (DK), vpravo nahoře Celle (DE), vlevo dole Osnabruck (DE) (zdroj: UA 2018).

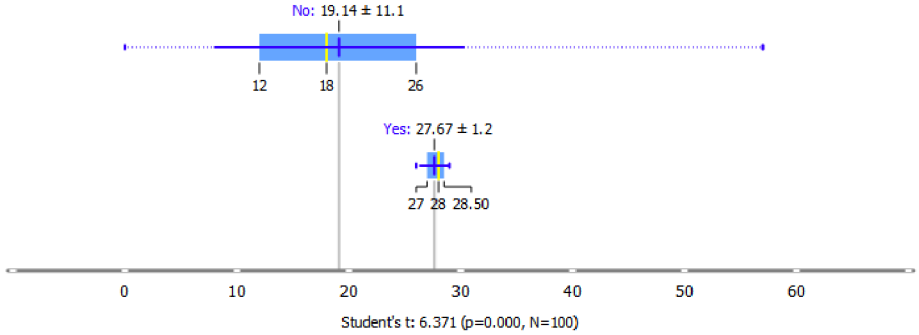
Okolí měst v rámci shluku je typické kombinací **orné půdy (21000), pastvin (23000) a lesů (31000)**, 

které se navzájem mísí a tvoří typickou trojbarevnou mozaiku, obdobou jako u shluku C7 Plauen – Gera. Podpora tohoto typu sousedství nabývá v rámci shluku průměrné hodnoty 33,33 % ± 6,6 (Obrázek 74).



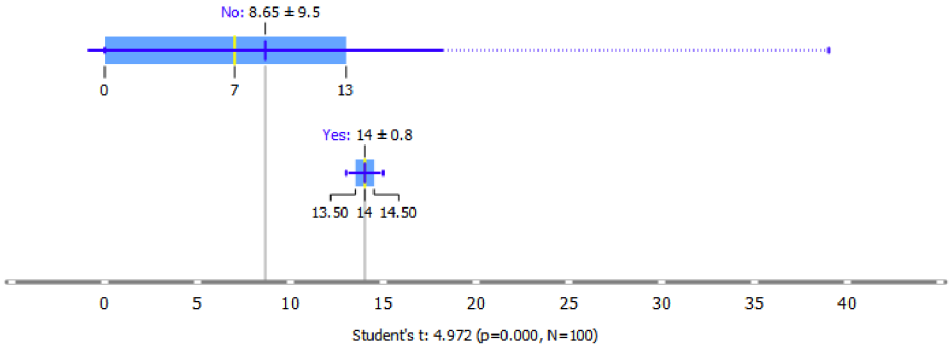
Obrázek 74 FS orné půdy (21000), pastvin (23000) a lesů (31000).

V centru města se vyskytuje v sousedství **nesouvislá hustá městská zástavba (11210), nesouvislá středně hustá městská zástavba (11220) a průmyslové, komerční a veřejné plochy (12100).** 12100 11210 11220 Průměrná podpora pravidla činila 27,67 % ± 1,2 (Obrázek 75).



Obrázek 75 FS nesouvislé husté městské zástavby (11210), nesouvislé středně husté městské zástavby (11220) a průmyslové, komerční a veřejné plochy (12100).

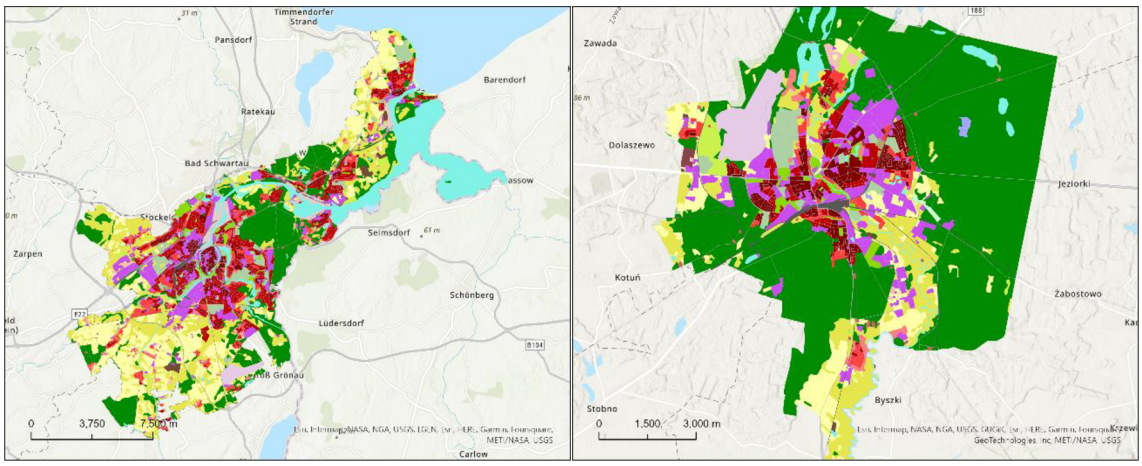
Druhou zajímavou kombinací v centru těchto měst je sousedství **nesouvislé středně husté městské zástavby (11220) a průmyslové, komerční a veřejné plochy (12100)** v blízkosti **městské zeleně (14100).** 12100 11220 14100 Průměrnou hodnotou podpory bylo 14 % ± 0,8 (Obrázek 76).



Obrázek 76 FS nesouvislé středně husté městské zástavby (11220), průmyslové, komerční a veřejné plochy (12100) a městské zeleně (14100).

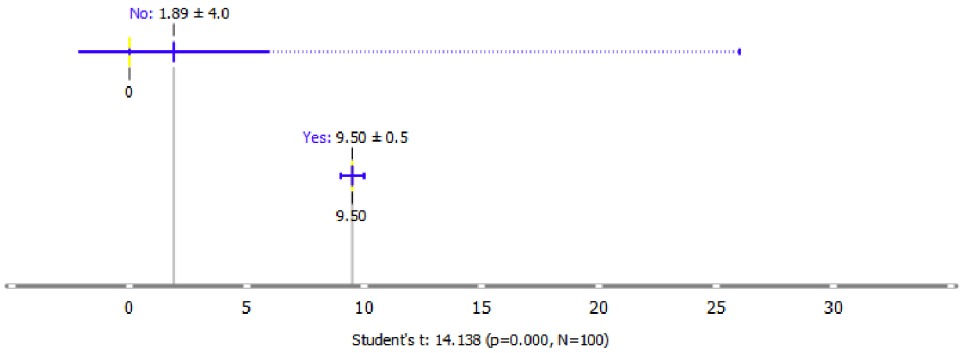
5.3.9 Lübeck (DE), Pila (PL) – C9

Lübeck je historické německé město s počtem obyvatel přibližně 220 000, nacházející se v severním Německu u Baltského moře. Je známé pro svůj historický střed s mnoha gotickými budovami, jako je například katedrála svatého Petra a stará radnice. Město bylo dříve důležitým obchodním centrem, zejména v oblasti obchodu s mořskou solí a jantarovým zbožím. Dnes je Lübeck turistickým cílem a má také důležitý přístav. **Piła** je menší město v severozápadním Polsku s počtem obyvatel přibližně 75 000. Nachází se v blízkosti řeky Gwda a je známé pro své parky a zelené plochy. Město má také mnoho kulturních institucí. Piła je důležitým průmyslovým centrem, zejména v oblasti dřevařského průmyslu.



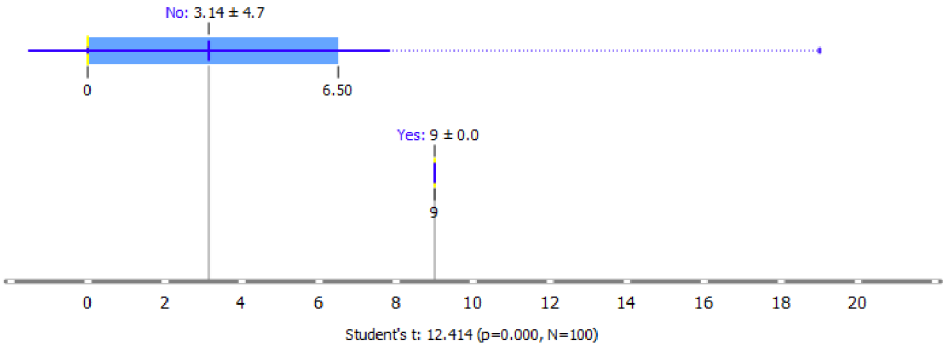
Obrázek 77 Náhled – vlevo Lübeck (DE), vpravo Piła (PL) (zdroj: UA 2018).

Okolí měst je typické kombinací **pastviny (23000), lesa (31000) a vodní plochy (50000).** 23000 31000 50000 Tato kombinace má průměrnou podporu 9,5 % ± 0,5. Takováto kombinace není u ostatních měst tak frekventovaná, jejich průměrná podpora činí 1,89 % ± 4,0. Studentův t-test s hodnotou 14,138 a hodnota p 0,000 naznačují statisticky významný rozdíl v průměrných hodnotách skupin (Obrázek 78).



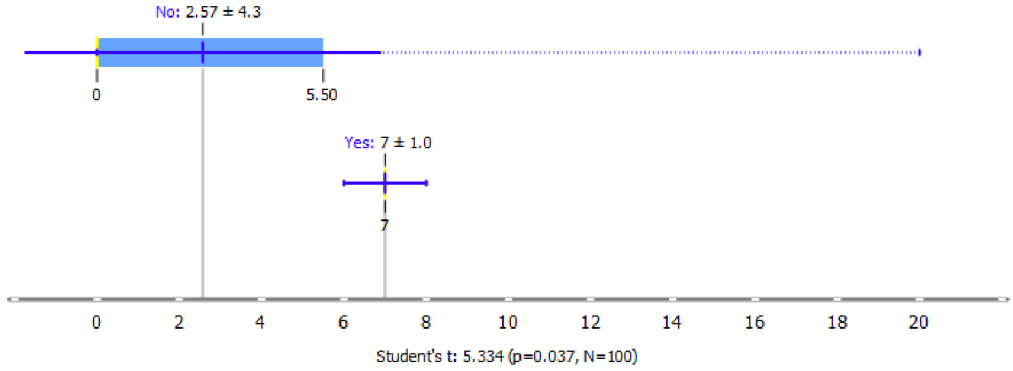
Obrázek 78 FS pastvin (23000), lesů (31000) a vodních ploch (50000).

Kombinací rurálních a urbánních využití území vznikla frekventovaná sada obsahující **nesouvislou hustou městskou zástavbu (11210), průmyslové, komerční a veřejné plochy (12100)** a **vodní plochy (50000)** . 12100 11210 50000 Tento typ sousednosti vzniká v důsledku přítomnosti vyššího množství vodních ploch. Celkově jsou obě města bohatá na frekventované sady sousednosti obsahující vodní plochy.



Obrázek 79 FS nesouvislá hustá městská zástavba (11210), průmyslové, komerční a veřejné plochy (12100) a vodní plochy (50000).

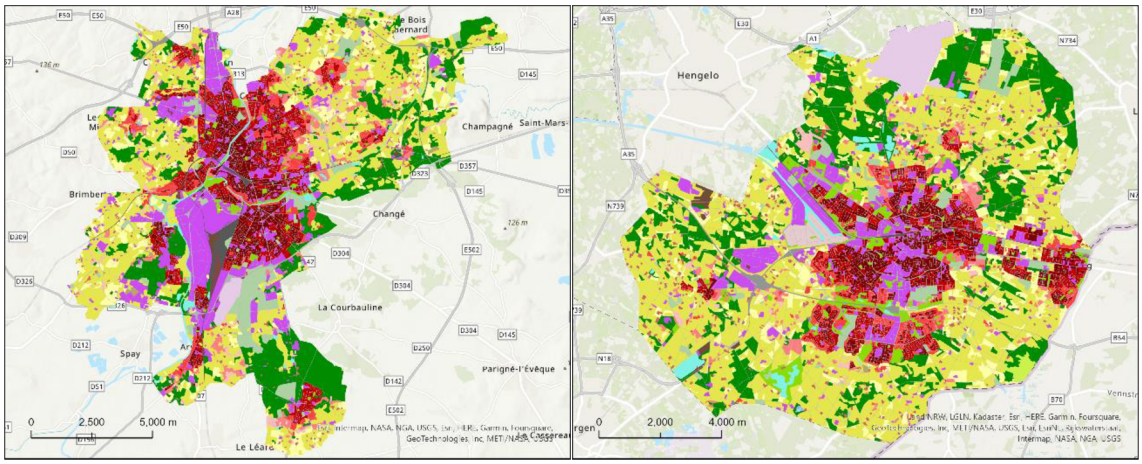
Další frekventovanou sadou, která má mírně vyšší průměrnou podporu oproti ostatním městům je **kombinace průmyslových, komerčních a veřejných ploch (12100), městské zeleně (14100) a vodních ploch (50000)**. 12100 50000 14100
Průměrná podpora v rámci shluku činí 7 % ± 1,0.



Obrázek 80 FS průmyslových, komerčních a veřejných ploch (12100), městské zeleně (14100) a vodních ploch (50000).

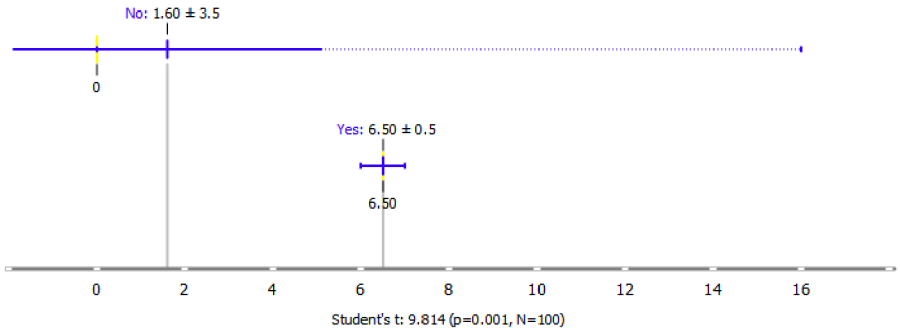
5.3.10 Le Mans (FR), Enschede (NL)– C10

Enschede je město na východě Nizozemska s populací přibližně 160 000 obyvatel. Je to významné průmyslové centrum a také kulturní a vzdělávací středisko, díky místní univerzitě a technické vysoké škole. **Le Mans** je město v západní Francii s populací přibližně 150 000 obyvatel. Je to významné průmyslové a obchodní centrum a také turistická destinace, díky slavnému závodu 24 hodin Le Mans a mnoha historickým památkám, jako je katedrála a hrad. Město má bohatou historii, která sahá až do římských dob.



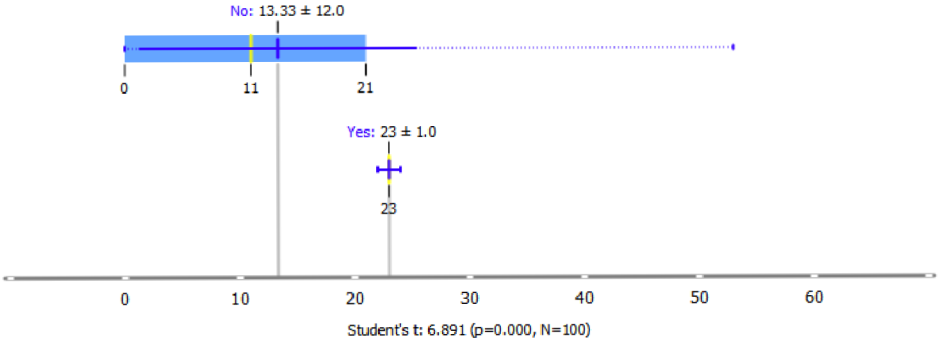
Obrázek 81 Náhled – vlevo Le Mans (FR), vpravo Enschede (NL) (zdroj: UA 2018).

Poměrně netypickou je kombinace sousednosti **průmyslových, komerčních a veřejných ploch (12100), izolovaných staveb (11300) a lesa (31000).** Tento druh sousedství 12100 31000 11300 se nachází v rurálních oblastech přilehajícím k lesům. Průměrná podpora sady činí 6,5 % ± 0,5 (Obrázek 82). Hodnota Studentova t-testu 9,814 ukazuje, že mezi průměry obou srovnávaných skupin je značný rozdíl, který je statisticky významný. Hodnota p rovna 0,001 naznačuje, že rozdíl mezi průměry je vysoce statisticky významný.

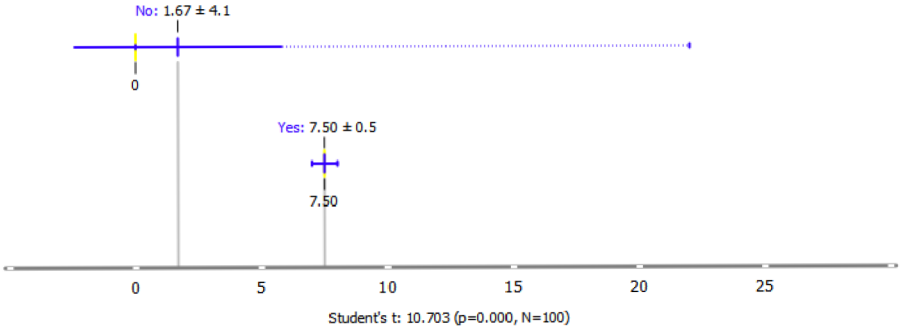


Obrázek 82 FS průmyslových, komerčních a veřejných ploch (12100), izolovaných staveb (11300) a lesa (31000).

Častou kombinací v rámci tohoto shluku byla kombinace **pastvin (23000) s nesouvislou málo hustou městskou zástavbou (11230)**, 23000 11230 popřípadě doplněná o **izolované stavby (11300).** 23000 11300 11230 U obou sad je průměrná hodnota podpory vyšší než u ostatních měst. Podpory činí 23 % a 7,5 % ± 0,5. U obou frekventovaných sad je rozdíl průměrů statistiky významný.



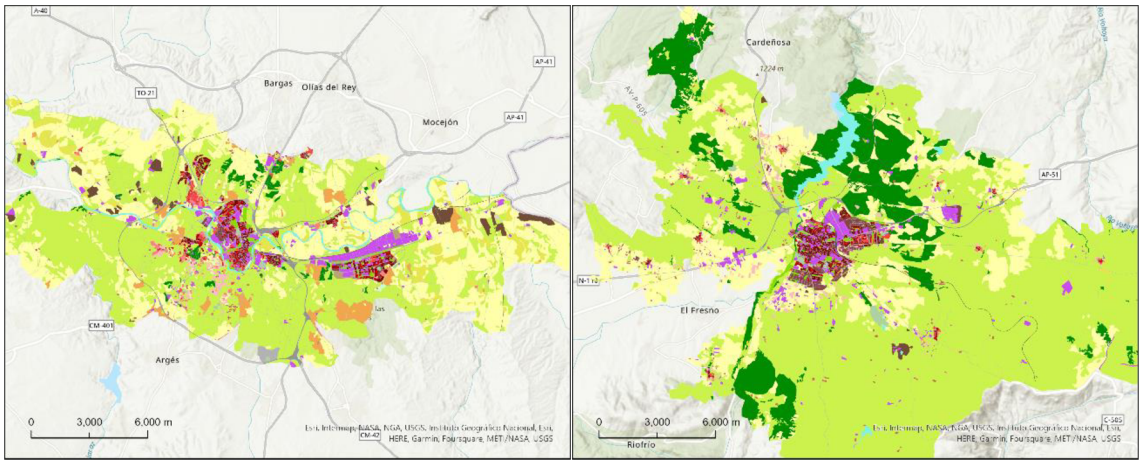
Obrázek 83 FS nesouvislé málo husté zástavby (11230) a pastvin (23000).



Obrázek 84 FS nesouvislé málo husté zástavby (11230), izolovaných staveb (11300) a pastvin (23000).

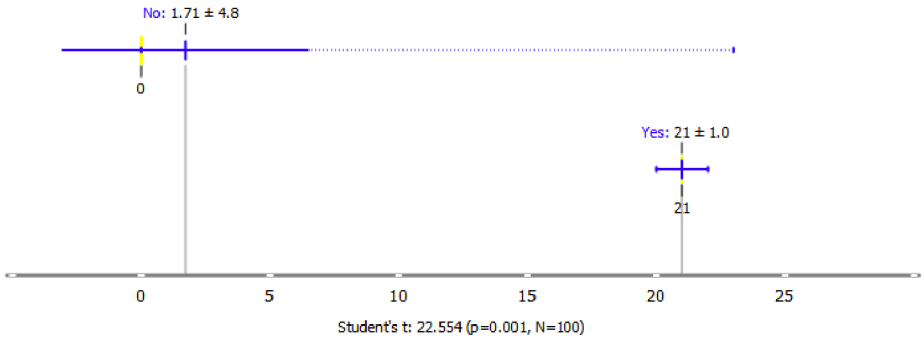
5.3.11 Toledo (ES), Avila (ES) – C11

Toledo a Ávila jsou dvě historická města nacházející se v centrální části Španělska. **Toledo** je hlavní město stejnojmenné provincie a má počet obyvatel přibližně 83 000. Nachází se na kopci nad řekou Tajo a je známé pro své historické památky. Toledo je také důležitým kulturním centrem Španělska a má mnoho festivalů a slavností. **Ávila** je menší město s počtem obyvatel přibližně 57 000 a nachází se asi 100 kilometrů západně od Madridu. Město je obehnané gotickými hradbami, které jsou jedním z nejvýznamnějších symbolů města.



Obrázek 85 Náhled – vlevo Toledo (ES), vpravo Avila (ES) (zdroj: UA 2018).

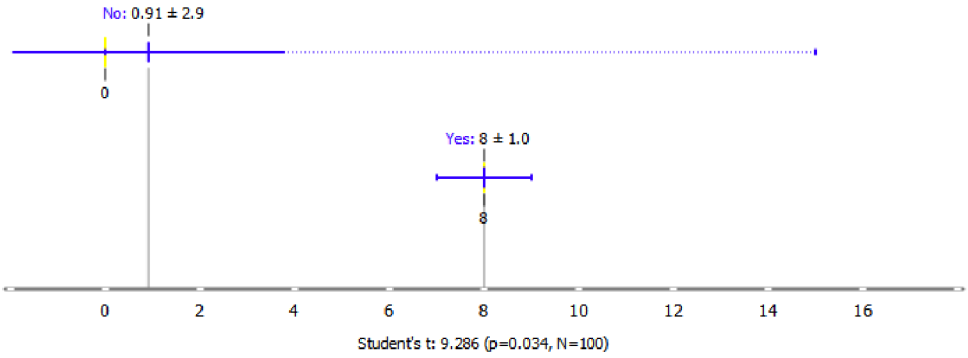
Okolí obou španělských měst je typicky pokryto územím s bylinnou vegetací (32000). První frekventovanou sadou sousednosti je kombinace **průmyslových, komerčních a veřejných ploch (12100), orné půdy (21000) a území s bylinnou vegetací (32000).** Tato sada se vyskytuje v rozšířeném centru města či mimo centrum. Průměr je roven 21 %± 1, zatímco u ostatních měst nabývá hodnot 1,17 %± 4,8. Studentův t-test s hodnotou 22,554 a p hodnotou 0,001 značí, že se jedná o významnou odlišnost v průměru (Obrázek 86).



Obrázek 86 FS průmyslových, komerčních a veřejných ploch (12100), orné půdy (21000) a území s bylinnou vegetací (32000).

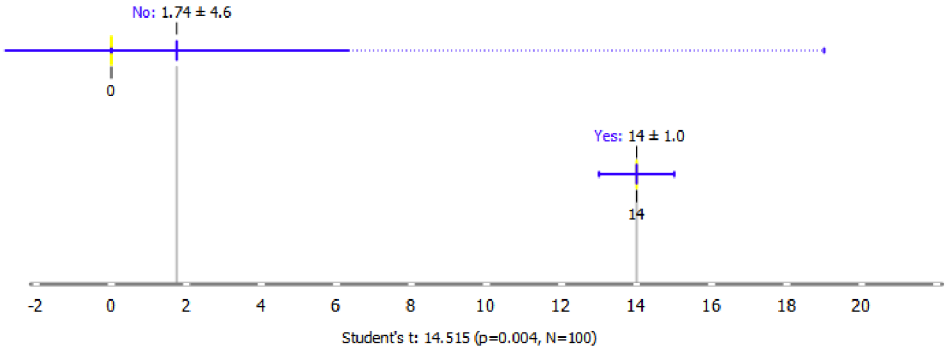
Další frekventovanou sadou je kombinace **souvislé husté městské zástavby (11100) průmyslových, komerčních a veřejných ploch (12100) a území s bylinnou vegetací (32000)**. Přítomnost této sady vypovídá o přechodu hustě zastavěného centra města do okolní vegetace. Průměr v této sadě s hodnotou 8 % ± 1 je

vyšší než průměr ostatních měst, který je 0,91 % ± 2,9. Hodnota Studentova t-testu je 9,286, p hodnota je 0,034 (Obrázek 87).

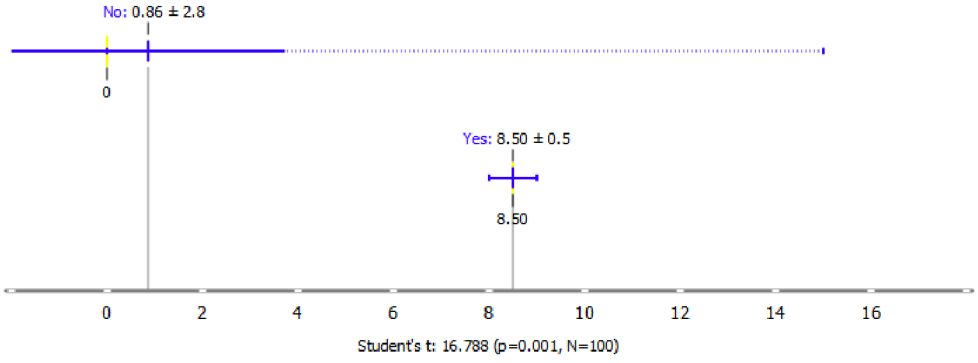


Obrázek 87 FS souvislé husté městské zástavby (11100), průmyslových, komerčních a veřejných ploch (12100) a území s bylinnou vegetací (32000)

Dalšími zajímavými frekventovanými sadami je sousednost **území s bylinnou vegetací (32000) s nesouvislou málo hustou městskou zástavbou (11230)** 32000 11230 **či s izolovanými stavbami (13400)** . 32000 11300 Toto sousedství nastává v oblastech mimo centrum města. V prvním případě se jedná o průměrnou podporu 14 % ± 1 (Obrázek 88). V druhém případě 8,5 % ± 0,5 (Obrázek 89). V obou případech je průměrná hodnota v ostatních městech nižší. Studentův t-test s hodnotami 14,515 a 16,788 s p hodnotami pod 0,005 značí o statistické významnosti odlišnosti průměrů.



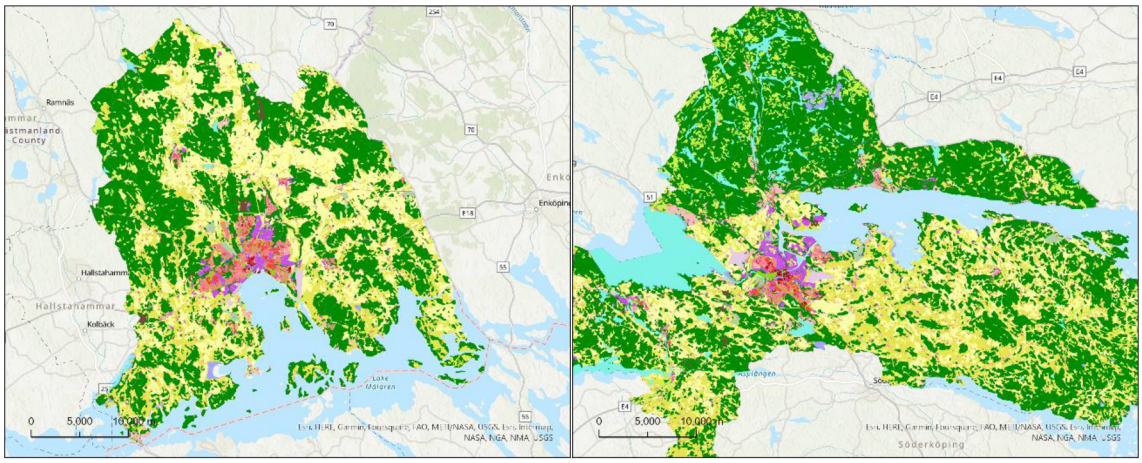
Obrázek 88 FS nesouvislé málo husté městské zástavby (11230) a území s bylinnou vegetací (32000).



Obrázek 89 FS izolovaných staveb (13400) a území s bylinnou vegetací (32000).

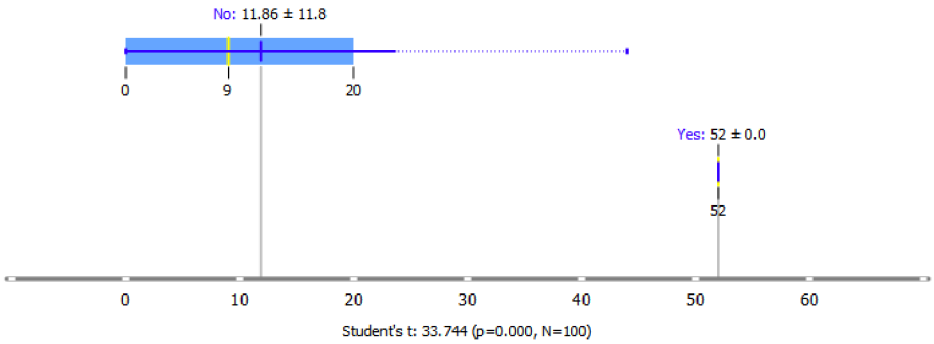
5.3.12 Vasteras (SE), Norrkoping (SE) – C12

Västerås a Norrköping jsou dvě města nacházející se ve Švédsku. **Västerås** je město v centrální části Švédska s počtem obyvatel přibližně 150 000. Nachází se na břehu jezera Mälaren a je jedním z nejstarších měst Švédska. Město má mnoho historických památek, jako je například katedrála, hrad a mnoho dalších kostelů a paláců. Västerås je také důležitým průmyslovým centrem, zejména v oblasti energetiky a elektrotechniky. Město má také mnoho parků a zelených ploch, což ho činí oblíbeným turistickým cílem. **Norrköping** je město s počtem obyvatel přibližně 100 000, které se nachází na východě Švédska u pobřeží Baltského moře. Město je známé pro své průmyslové dědictví, zejména v oblasti textilního průmyslu. Město je také důležitým kulturním centrem Švédska a má mnoho festivalů a kulturních akcí.



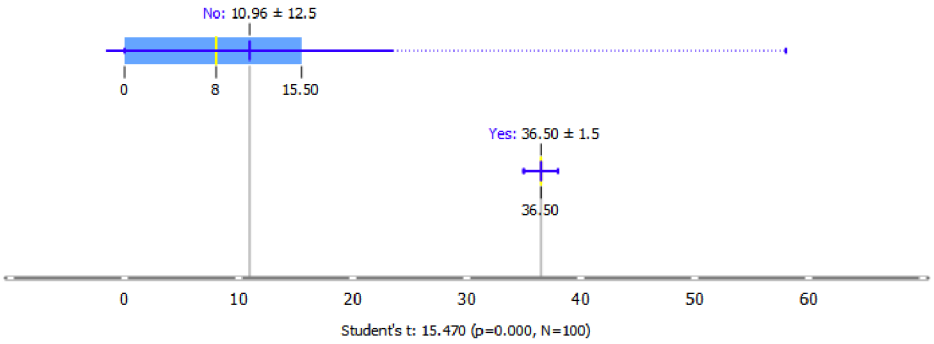
Obrázek 90 Náhled – vlevo Vasteras (SE), vpravo Norrkoping (SE) (zdroj: UA 2018).

Při bližším pohledu na urbánní jádra měst v rámci shluku si můžeme povšimnout velké rurální části. Problematice vymezených území je věnována část v rámci kapitoly Diskuze. Významně odlišnou je tedy frekventovaná sada sousednosti **orné půdy (21000), pastvin (23000) a lesů (31000)** 31000 23000 21000 s průměrnou podporou 52 %. Průměrná podpora napříč ostatními městy je 11,96 ± 11,8, tedy výrazně nižší. Při pohledu na krabicový graf můžeme u této frekventované sady vidět, že žádné další město nenabývá vyšší hodnoty pravidla než přibližně 45 %. Hodnota Studentova t-testu 33,744 a hodnota p 0,000 indikuje významnou statistickou významnost odlišnosti průměrů.



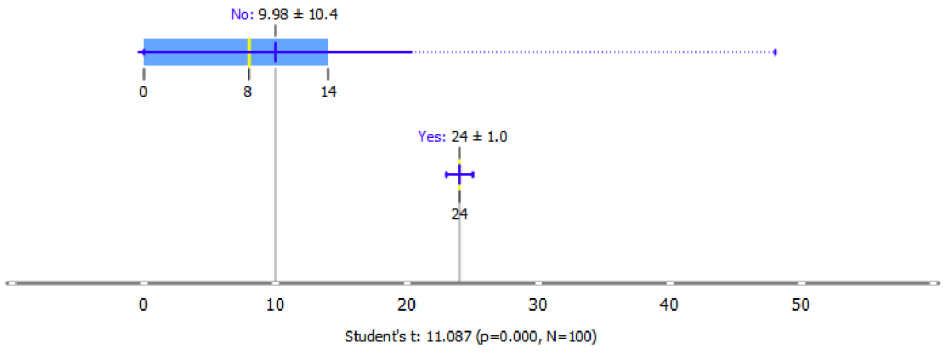
Obrázek 91 FS orné půdy (21000), pastvin (23000) a lesů (31000).

Další významně odlišnou frekventovanou sadou je kombinace **izolovaných staveb (11300) a pastvin (23000)**, 23000 11300 která se nachází rozesetá v širokém okolí měst v rámci shluku. Průměrná podpora pravidla činí 36,5 % ± 1,5 (Obrázek 92). Obdobně četnou podporu mají i sady izolovaných staveb s lesy (31000) a ornou půdou (21000).



Obrázek 92 FS izolovaných staveb (11300) a pastvin (23000).

Co se týče zástavby, typickou je kombinace sousedství **nesouvislé velmi málo husté zástavby (11240) a pastviny (23000)**. 23000 11240 Tento druh sousedství se nachází při okrajích zastavěného území vlastních měst. Průměrná podpora frekventované sady činí 24 % ± 1,0 (Obrázek 93) a na základě hodnoty Studentova t-testu a hodnoty p je jedná o statisticky významný rozdíl oproti průměru ostatních měst (9,98 % ± 10,4). Obdobně jako u předchozího pravidla jsou rovněž frekventované kombinace s lesy (31000) a ornou půdou (21000).

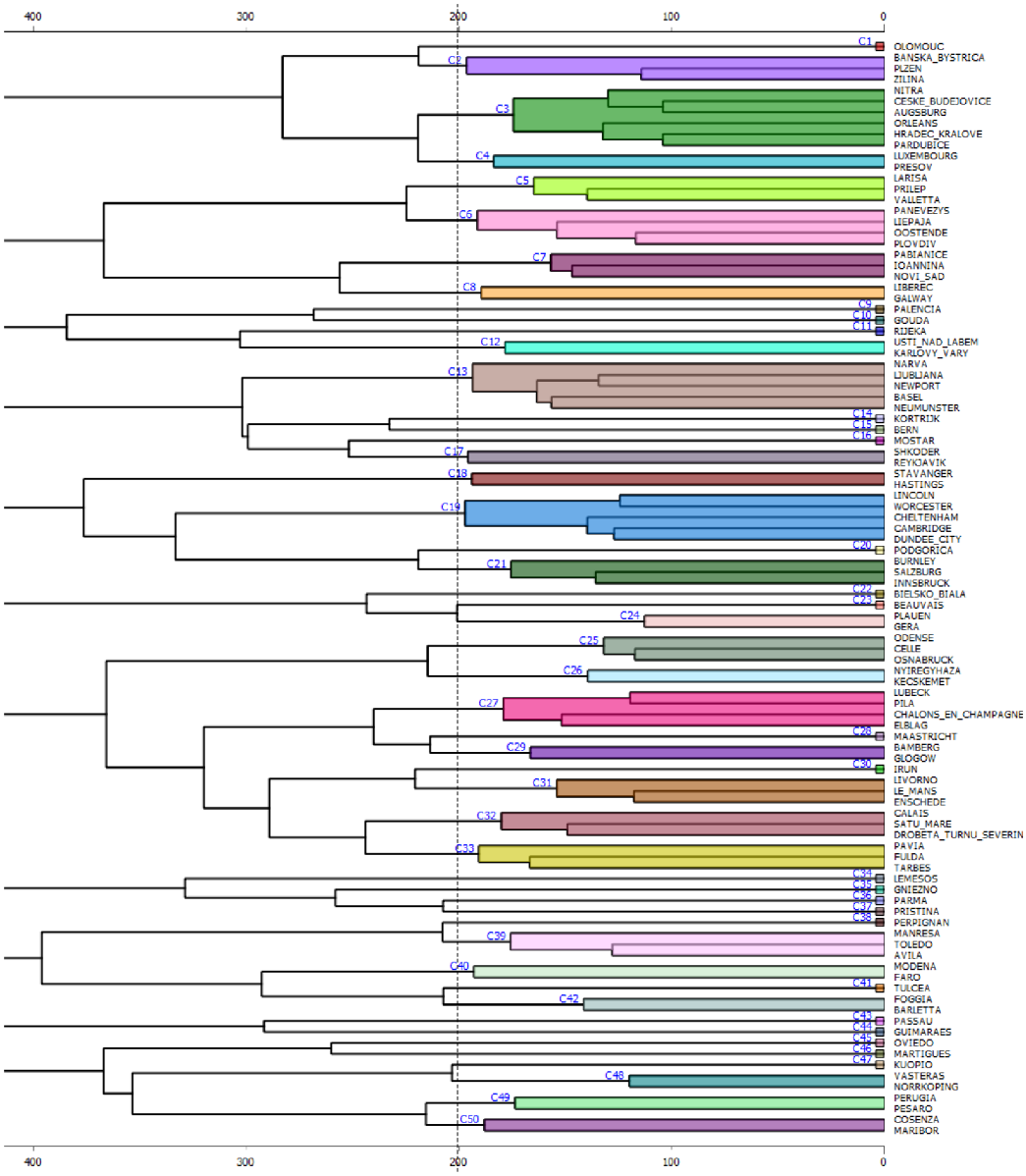


Obrázek 93 FS nesouvislé velmi málo husté zástavby (11240) a pastviny (23000).

5.4 Porovnání s výsledky studie (Dobesova 2020)

V rámci vstupních dat bylo zahrnuto 22 měst ze studie Experiment in Finding Look-Alike European Cities Using Urban Atlas Data (Dobesova 2020). Jednalo se o dvojice měst (Tabulka 21), která byla identifikována jako podobná na základě zkoumání dat využití území datasetu Urban Atlas s využitím metody k-Nearest Neighbor nad feature vektorem získaným z neuronové sítě Painters (Kaggle 2016).

Studie v rámci této práce měla za vstup významně menší datovou sadu. Bylo počítáno s výběrem 100 měst namísto kompletní datové sady v rámci porovnávané studie (přibližně 800 měst). Toto zmenšení datové sady mělo pravděpodobně vliv na nalezené podobnosti. Na druhou stranu byly zahrnuty všechny zjištěné dvojice v rámci porovnávané studie (24 měst). Cílem porovnání bylo potvrdit, či vyvrátit zjištěné podobnosti. Dalším významným rozdílem oproti původní studii je vymezené území. Srovnávaná studie využívala kruhové výseče se středem v centru města s pouze minimálním přesahem do okolní krajiny. Tato práce pracuje s vymezením urbánních jader v rámci datasetu Urban Atlas 2018. Blíže se problematice vymezeného území věnuje kapitola Diskuze.



Obrázek 94 Hierarchické shlukování s prahovou vzdáleností 200.

Tabulka 21 Dvojice podobných měst ze studie (Dobesova 2020) – červeně zjištěna stejná podobnost.

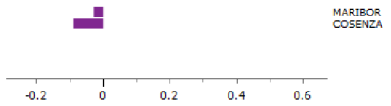
Maribor	Bern
Le Mans	**Enschede**
České Budějovice	**Hradec Králové**
Modena	Parma
Plovdiv	Perpignan
Bielsko-Biala	Basel
Perugia	Plauen
Guimaraes	Osnabruck
Ljubljana	Lübeck
Enschede	Oviendo
Glogow	Maastricht

Podobnost je vyhodnocována za základě **hierarchického shlukování na základě euklidovské vzdálenosti a Wardovy metody**. V rámci dendrogramu jsou jako podobná města považována ta, která se společně nacházejí v rámci shluku s **prahovou vzdáleností 200**. Tato vzdálenost je větší, než při zkoumání v rámci studie 5.3 Případová studie – evropská města, kde byla použita vzdálenost 125. Rovněž je zkoumána kvalita shlukování pomocí grafů siluety a jednotlivá shlukování jsou hodnocena.

Při porovnání měst mezi studiemi byla zjištěna stejná podobnost mezi městy **Le Mans ve Francii a Enschede v Nizozemsku** a **Českými Budějovicemi a Hradcem Králové v České republice.** Důvody podobnosti byly dříve představeny v rámci kapitoly 5.3 Případová studie – evropská města.

5.4.1 Maribor (SI) – Bern (CH)

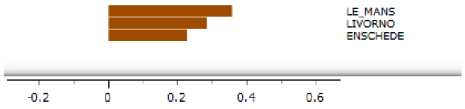
Podobnost Mariboru ve Slovinsku a Bernu ve Švýcarsku se na základě hierarchického shlukování nepotvrdila. Nejpodobnějším městem Mariboru na základě analýzy sousednosti využití území do 100 metrů byla Cosenza v Itálii. Hodnota grafu siluety 0 pro Maribor naznačuje nejednoznačnost přiřazení datového bodu ke shluku, zatímco záporná hodnota, -0,1 pro Cosenzu, naznačuje, že datový bod nemusí patřit do přiřazeného shluku a může být nutné jej znovu přiřadit nebo dále prozkoumat. Na základě hodnocení konzistence shluku lze říci, že se nejedná o významně podobná města. Bern se v dané prahové vzdálenosti nestal součástí žádného shluku.



Obrázek 95 Graf siluety shluku Maribor a Cosenza.

5.4.2 Le Mans (FR) – Enschede (NL)

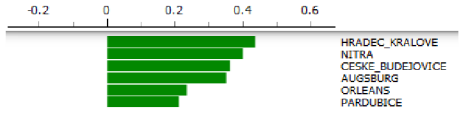
Podobnost Le Mans ve Francii a Enschede v Nizozemsku se potvrdila. V rámci nastavení prahové vzdálenosti při hierarchickém shlukování do skupiny podobných měst přibylo Livorno v Itálii. Hodnoty grafu siluety (Obrázek 96) 0,2, 0,3 nebo 0,4 obecně znamenají, že datové body jsou poměrně dobře shlukovány a mají poměrně vysoký stupeň podobnosti s ostatními datovými body v rámci přiřazeného shluku. Čím větší je hodnota koeficientu siluety, tím lépe daný datový bod odpovídá shluku.



Obrázek 96 Graf siluety shluku Le Mans, Livorno a Enschede.

5.4.3 České Budějovice (CZ) – Hradec Králové (CZ)

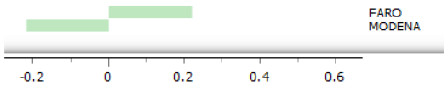
Tato dvě česká krajská města v rámci detailního hierarchického shlukování skončila ve dvou samostatných shlucích, které se ale v rámci dendrogramu vyskytovaly vedle sebe. Následně došlo ke sloučení těchto dvou shluků [Nitra (SK), České Budějovice (CZ), Augsburg (DE) – C2] a [Orleans (FR), Hradec Králové (CZ), Pardubice (CZ) – C3]. Města lze prohlásit za podobná. Hodnoty grafu siluety (Obrázek 97) poukazují na poměrně vysoká stupně podobnosti v rámci shluku.



Obrázek 97 Graf siluety shluku Hradec Králové, Nitra, České Budějovice, Augsburg, Orleans a Pardubice.

5.4.4 Modena (IT) – Parma (IT)

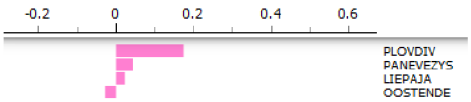
Podobnost Modeny a Parmy se aplikací metody analýzy frekventovaných sad sousednosti nepotvrdila. Parma se při prahové vzdálenosti 200 neslučovala do žádného shluku. Modena tvořila dvojici s městem Faro v Portugalsku. Kvalita shluku nebyla dobrá. Záporné i kladné hodnoty grafu siluety ve stejném shluku naznačují, že shluk nemusí být dobře oddělen od ostatních shluků a že některé datové body v rámci shluku mohou být chybně klasifikovány nebo nejednoznačné.



Obrázek 98 Graf siluety shluku Faro a Modena.

5.4.5 Plovdiv (BG) – Perpignan (IT)

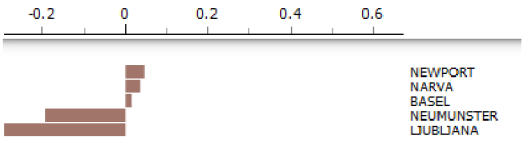
Podobnost Plovdivu v Bulharsku a Perpignanu v Itálii se nepotvrdila. Perpignan v rámci prahové vzdálenosti dendrogramu nevstupoval do žádného shluku. Plovdiv tvořil shluk s městy Panevezys v Litvě, Liepaja v Lotyšsku a Oostende v Belgii. Kvalita shluku je suboptimální se směsí kladných a záporných hodnot koeficientu siluety (Obrázek 99), což svědčí o nesprávné klasifikaci nebo nejednoznačnosti některých datových bodů.



Obrázek 99 Graf siluety shluku Panevezys, Liepaja, Oostende a Plovdiv.

5.4.6 Bielsko-Biala (PL) – Basel (CH)

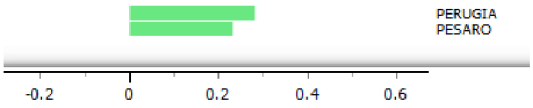
Podobnost dvojice měst se nepotvrdila. Bielsko-Biala netvořila při tomto nastavení žádný shluk s ostatními městy. Basel se vyskytovala v rámci shluku Narva, Ljubjana, Newport. Kvalita shluků není optimální (Obrázek 100), protože záporné hodnoty koeficientu siluety (-0,2) naznačují, že dva z datových bodů mohou být špatně klasifikovány nebo více podobné datovým bodům v jiných shlucích.



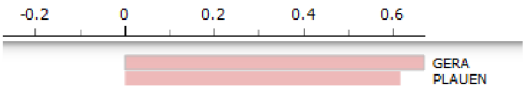
Obrázek 100 Graf siluety shluku Narva, Ljubjana, Newport a Basel.

5.4.7 Perugia (IT) – Plauen (DE)

Města Perugie v Itálii a Plauen v Německu netvořila společný shluk v rámci nastavení hierarchického shlukování. Perugia tvořila shluk s městem Pesaro rovněž v Itálii. Kvalita shluku je střední, protože oba datové body mají kladné hodnoty koeficientu siluety (Obrázek 101) (0,2 a 0,3), což naznačuje, že jsou poměrně dobře shlukovány a mají určitou podobnost mezi sebou v rámci přiřazeného shluku. Plauen tvořilo shluk s městem Gera rovněž v Německu. Kvalita shluku je dobrá, protože oba datové body mají vysoké kladné hodnoty koeficientu siluety (Obrázek 102) (0,6 a 0,7), což znamená, že jsou dobře shlukovány a mají vysoký stupeň podobnosti mezi sebou v rámci přiřazeného shluku.



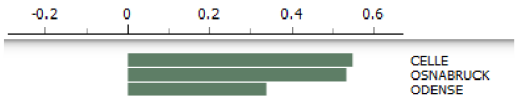
Obrázek 101 Graf siluety shluku Perugia a Pesaro.



Obrázek 102 Graf siluety shluku Gera a Plauen.

5.4.8 Guimaraes (PT) – Osnabruck (DE)

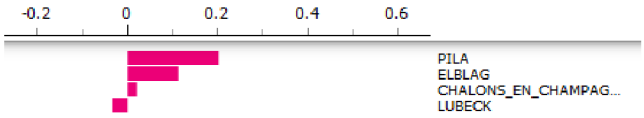
Podobnost měst Guimares v Portugalsku a Osnabruck v Německu se nepotvrdila. Guimaraes se v rámci nastavených prahových hodnot neshlukovalo s žádným městem. Osnabruck tvořil shluk s městy Odense v Dánsku a Celle v Německu (5.3.8 Odense (DK), Celle (DE) Osnabruck (DE) – C8). Kvalita shluku je střední až dobrá, protože dva z datových bodů mají vysoké kladné hodnoty koeficientu siluety (Obrázek 103) (0,55), což znamená, že jsou dobře shlukovány a mají vysokou míru podobnosti s ostatními datovými body v rámci přiřazeného shluku. Odense má o něco nižší hodnotu koeficientu siluety (0,3), což naznačuje, že je stále poměrně dobře shlukován, ale může mít o něco nižší stupeň podobnosti s ostatními datovými body v rámci shluku.



Obrázek 103 Graf siluety shluku Odense, Celle a Osnabruck.

5.4.9 Ljubjana (SI) – Lübeck (DE)

Podobnost měst Ljubjana a Lübeck se při hierarchickém shlukování neprokázala. Ljubjana se vyskytovala v již popsaném shluku Narva, Ljubjana, Newport a Basel se špatnou kvalitou shlukování. Lübeck se nacházel v rámci shluku s městy Pila v Polsku, Chalons en Champagne ve Francii a Elblag v Polsku. Kvalita shluků je střední až špatná, protože kladné hodnoty koeficientu siluety (0,2 a 0,1) naznačují, že dva z datových bodů (Pila, Elblag) jsou poměrně dobře shlukovány a mají určitou podobnost s ostatními datovými body v rámci přiřazeného shluku, ale záporná hodnota koeficientu siluety (-0,1) naznačuje, že jeden datový bod (Lübeck) může být špatně klasifikován nebo více podobný datovým bodům v jiných shlucích. Přítomnost záporné hodnoty koeficientu siluety naznačuje, že daný shluk nemusí být dobře oddělen od ostatních shluků.



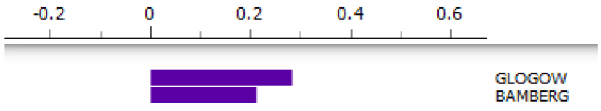
Obrázek 104 Graf siluety shluku Lübeck, Pila, Chalons en Champagne a Elglag.

5.4.10 Enschede (NL) – Oviendo (ES)

Podobnost měst Enschede a Oviendo se neprokázala. Oviendo ve Španělsku se nenacházelo v rámci žádného shluku. Naopak Enschede se nacházelo v rámci shluku Le Mans, Livorno, Enschede blíže popsaného v rámci podkapitoly 5.4.2 Le Mans (FR) – Enschede.

5.4.11 Glogow (PL) – Maastricht (NL)

Podobnost na základě analýzy sousednosti využití území a frekventovaných sad se u měst Glogow v Polsku a Maastricht v Nizozemsku neprokázala. Maastricht se nenacházel v rámci žádného shluku. Město Glogow tvořilo shluk s městem Bamberg v Německu. Kvalita shluku je střední až špatná, protože oba datové body mají kladné hodnoty koeficientu siluety Obrázek 105 Graf siluety shluku Glogow a Bamberg.(Obrázek 105) (0,3 a 0,2), které jsou však relativně nízké, což naznačuje, že jsou pouze středně dobře shlukovány a mají určitý stupeň podobnosti mezi sebou v rámci přiřazeného shluku.



Obrázek 105 Graf siluety shluku Glogow a Bamberg.

6 DISKUZE

Práce aplikuje data miningovou techniku dolování frekventovaných sad na prostorová data. Frekventované sady se převážně používají při analyzování transakčních sad typu nákupní košík. V rámci rešerše nebyly nalezeny žádné studie, které by techniku aplikovaly právě na prostorová data. Hlavní výzvou nasazení frekventovaných sad byla podstata prostorových dat. Podstatou metody frekventovaných sad je snaha říci, co se vyskytuje, s čím a s jakou podporou a v tomto řešeném případě přenesené do prostoru, kdy současný výskyt je zaznamenán jako sousednost typů landuse. Práce aplikuje metodu na transakční data sousednosti polygonů využití území v rámci datasetu Urban Atlas 2018.

Příkladem aplikace analýzy frekventovaných sad může být aplikace na **data Twitteru.** Metoda umožňuje získat shrnutí nejčastěji používaných slov nebo frází v daném datasetu tweetů. Tento proces zahrnuje extrakci slov nebo frází z tweetů, odstranění **stop slov** a následné spočítání jejich výskytů. Stop slovo je slovo, které se vyskytuje velmi často v jazyce a obvykle nese málo významu pro porozumění textu. Tato slova jsou obvykle vyloučena z analýzy textu, protože neodrážejí konkrétní obsah nebo téma, na které se text zaměřuje. Příklady stop slov v angličtině zahrnují slova jako "the", "and", "of", "a", "an", "in" a "to". Odstranění stop slov z textových dat může pomoci zlepšit kvalitu analýzy a snížit množství šumu v datových sadách. Když se přeneseme na aplikaci této metody na transakční data sousednosti využití území, můžeme za „stop slovo“ považovat sousedství se silnicí v rámci města. **Silnice se v rámci evropských měst vyskytují ve velké míře** a podpora pravidel s nimi není pro určení podobnosti klíčová, protože podpora dosahuje ve všech městech vysokých hodnot. V budoucích aplikacích by tedy bylo možné dopředu vybrat pouze některé typy využití území.

Samotná metoda **nepracuje s četností jednotlivých prvků v rámci transakce sousednosti**. Při generování transakčních dat sousedství pomocí implementovaného nástroje tak stačí pro zanesení údaje o sousednosti pouze malý překryv a na druhou stranu významně velký překryv bude také odpovídat jednomu záznamu v rámci transakce. Tento nedostatek je částečně odbourán velkým množství transakčních dat v rámci datasetu, a tak tento druh nepřesnosti se částečně odstraní velkým množstvím záznamů.

Další výzvou při aplikaci metody na data využití území je **tematická obsáhlost některých kategorií landuse vstupních dat**. Například kategorie 12100 zahrnuje průmyslové, komerční, veřejné a vojenské plochy. Tato kategorie je široká a znemožňuje rozlišování mezi skutečným funkčním využitím území. Komerční plochy (např. supermarkety) využívají obyvatelé města zcela jinak než vojenské plochy a plní i jiný účel. Zde by bylo do budoucna možné nějakým způsobem zpřesnit některé široce definované kategorie použitím zpřesňujících dat. To by bylo ale poměrně komplikované napříč celým datasetem, který obsahuje převážnou většinu evropských států.

Pro detailnější vyhodnocení měst a jejich různorodosti je problematické i **vlastní rozdělení kategorií landuse Urban Atlas**. Prvních šest kategorií – 11100 až 11240 jsou kategorie Urban Fabric – městské zástavby s různou úrovní hustoty zástavby. Tyto kategorie vyjadřují pouze strukturu, či hustotu zástavby, nikoliv její konkrétní využití (landuse), které může být i např. smíšené pro bydlení a obchod. Ostatní kategorie naopak vyjadřují pouze využití (landuse), nikoliv strukturu (hustotu) jako jsou kategorie zástavby. Z frekventovaných sad tudíž nelze usuzovat na detailnější fungování města.

Nicméně lze vyslovit myšlenku, že víceprvkové frekventované sady (4 a více prvků) s vyšší hodnotou podpory (10 až 30 %) poukazují na různorodější **mix landuse**, který je vnímán urbanisty jako vhodnější pro život obyvatel měst. Přítomnost jedno, dvou a tří prvkových frekventovaných sad s vysokou podporou poukazuje na homogenní oblasti ve

městě s nízkým landuse mixem. Zejména frekventované sady s kombinací kategorie 12100 (průmyslové, komerční a veřejné plochy) a 12220 (silnice) a vysokou podporou lze považovat za města orientovaná více na ekonomické funkce než funkce správní, kulturní či pro bydlení. Tyto myšlenky, kdy určité kombinace frekventovaných sad s určitou podporou můžou indikovat různé funkční určení měst, či jejich historii a vývoj městského plánování, mají potenciál rozvinutí v dalším výzkumu.

Významným problémem vstupních dat je **nekonzistentnost definice městského jádra** (Urban Core) napříč jednotlivými státy datasetu. V rámci práce je striktně pracováno s administrativní jednotkou hranice městského jádra definovaného v rámci datasetu Například Albánie a Severní Makedonie používají k vymezení městského jádra čtvercovou síť o straně 1 km. Naopak Norsko a Švédsko zahrnují významně velké oblasti lesů a pastvin v okolí měst. Takovéto město má následně vysoké podpory rurálních typů frekventovaných sad. Problém by šel částečně odstranit použitím např. kruhové výseče z práce (Janoušek 2017), popřípadě vlastní harmonizací rozsahu vymezeného území.

Jako problémové se ukázalo **vyhodnocení frekventovaných sad**. Na procentuální podporu jednotlivých sad lze použít shlukovací metody typu hierarchického shlukování, ale nalezení konkrétních frekventovaných sad, které způsobují dané shlukování je obtížné. V rámci práce byly takové sady hledány manuálně, a to na základě vizuálního procházení měst v rámci shluků v mapě. Dále je potřeba zmínit vliv celkového počtu unikátních frekventovaných sad. Sumarizační matice v případové studii (5.3 Případová studie – evropská města) měla 3460 unikátních frekventovaných sad (tj. sloupců), kdy řada obsahovala nulovou hodnotu podpory. Jak bylo zjištěno v kapitole (4.4 Nastavení vzdálenosti nástroje) vybraná vzdálenost má významný vliv na celkový počet frekventovaných sad.

ZÁVĚR

Cílem práce bylo srovnání evropských měst na základě frekventovaných sad vyjadřující sousednost různých landuse (využití území) v městech. Dalším cílem byla realizace programu na přípravu kategoriálních nebo dichotomických dat sousednosti polygonů nutných pro generování frekventovaných sad. Na základě zjištěných frekventovaných sad bylo cílem popsat charakter měst (1), jejich porovnání a nalezení podobných měst (2). Zdrojová data využití území byla použita z Copernicus Urban Atlas 2018.

Celkově lze říci, že práce **implementuje netradiční metodu**, **běžně nepoužívanou v rámci geoinformatiky**. Samotné případové studie měly za cíl primárně ověřit možnosti využití metody. Primárním výsledkem práce je nalezení postupů pro aplikování metody.

Teoretická část práce se zaobírala obecnými koncepty Data Miningu, popisem algoritmů pro výpočet frekventovaných sad a popisem studií, které více či méně implementují metodu frekventovaných sad na prostorová data.

Prvním mezivýsledkem práce byla **implementace nástroje pro generování transakčních dat sousednosti využití území** blíže představeného v rámci kapitoly 4.3.3 **SearchDistinctLanduse_SpatialJoin**. Nástroj v softwaru ArcGIS Pro zjistí všechny kódy využití landuse sousedních polygonů až do uživatelem definované vzdálenosti. Výstupem nástroje jsou transakční data sousednosti ve dvou formátech výstupních dat. Prvním je textový soubor obsahující kategoriální data – kódy landuse dle Urban Atlas. Textový soubor lze dále zpracovávat v programu SPMF nebo programem negFI. Druhým výstupem je soubor MS Excel obsahující dichotomická data, kde hodnoty 0 a 1 nesou informaci o přítomnosti nebo nepřítomnosti každého z 27 kódu landuse v sousedství. Formát MS Excel je vhodný pro následné použití v programu Orange.

V rámci kapitoly (5.1 **Popis vybraných měst**) byla představena možná intepretace vypočítaných frekventovaných sad na příkladu měst Cheltenham (UK) a Prešov (SK). Následně byly zpracovány 3 případové studie s cílem otestovat aplikovatelnost metody na prostorová data.

Cílem 1. případové studie (5.2 **Případová studie – česká města**) bylo poskytnout souhrnný popis českých měst v rámci sady UA 2018, která zahrnuje celkem 15 měst. Tato skupina měst zahrnuje krajská města České republiky a dále také města Most a Chomutov. Pro vybraná česká města byly identifikovány signifikantní frekventované sady. Byly nalezeny kombinace využití území, které jsou typické pro města v České republice. Provedení podobných analýz pro jednotlivé národní státy v rámci daného datového souboru by umožnilo porovnat a identifikovat podobnosti a rozdíly mezi těmito státy.

Cílem 2. případové studie (5.3 **Případová studie – evropská města**) bylo prozkoumat 100 vybraných evropských měst a na základě frekventovaných sad sousednosti využití území nalézt podobná města. Pomocí hierarchického shlukování bylo nalezeno 12 skupin o 27 městech, jejichž charakter byl popsán v podkapitolách 5.3.1 až 5.3.12.

Cílem 3. případové studie (5.4 **Porovnání s výsledky studie (Dobesova 2020)**) bylo porovnat získané skupiny s výsledky jiné studie. Jako vstupní data bylo použito 22 měst ze studie (Dobesova, 2020). Tato studie identifikovala dvojice měst, která jsou si podobná na základě využití území datasetu Urban Atlas pomocí metody k-Nearest Neighbor nad feature vektorem získaným z neuronové sítě Painters (Kaggle, 2016). Byla potvrzena podobnost dvou dvojic měst Le Mans – Enschede a České Budějovice – Hradec Králové.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

AGRAWAL, Rakesh a Ramakrishnan SRIKANT, 1994. Fast Algorithms for Mining Association Rules in Large Databases. In: *Proceedings of the 20th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., s. 487–499. VLDB '94. ISBN 1-55860-153-8.

ARYABARZAN, Nader, Behrouz MINAEI-BIDGOLI a Mohammad TESHNEHLAB, 2018. negFIN: An efficient algorithm for fast mining frequent itemsets. *Expert Systems with Applications* [online]. 105, 129–143. ISSN 0957-4174. Dostupné z: doi:10.1016/j.eswa.2018.03.041

BERKA, Petr, 2003. *Dobývání znalostí z databází*. Vyd. 1. Praha: Academia. ISBN 978-80-200-1062-9.

BERRY, Michael J. A. a Gordon LINOFF, 2004. *Data mining techniques: for marketing, sales, and customer relationship management*. 2nd ed. Indianapolis, Ind: Wiley Pub. ISBN 978-0-471-47064-9.

BOEING, Geoff, 2018. Measuring the complexity of urban form and design. *URBAN DESIGN International* [online]. 23(4), 281–292. ISSN 1357-5317, 1468-4519. Dostupné z: doi:10.1057/s41289-018-0072-1

BOEING, Geoff, 2019. Urban spatial order: street network orientation, configuration, and entropy. *Applied Network Science* [online]. 4(1), 67. ISSN 2364-8228. Dostupné z: doi:10.1007/s41109-019-0189-1

BOEING, Geoff, 2020. A multi-scale analysis of 27,000 urban street networks: Every US city, town, urbanized area, and Zillow neighborhood. *Environment and Planning B: Urban Analytics and City Science* [online]. 47(4), 590–608. ISSN 2399-8083, 2399-8091. Dostupné z: doi:10.1177/2399808318784595

BUČKOVÁ, Simona, 2022. *Aplikace vyhledávání kolokačních vzorů na prostorová data* [online]. Olomouc [vid. 2023-02-25]. Univerzita Palackého v Olomouci, Přírodovědecká fakulta. Dostupné z: https://theses.cz/id/10j2jw/?zpet=%2Fvyhledavani%2F%3Fsearch%3Daplikace%20vyhled%C3%A1v%C3%A1n%C3%AD%20koloka%C4%8Dn%C3%ADch%26start%3D1;isshlret=Aplikace%3Bvyhled%C3%A1v%C3%A1n%C3%AD%3Bkoloka%C4%8Dn%C3%ADch%3B

BURIAN, Jaroslav, Lenka ZAJICKOVA a Igor IVAN, 2016. Analýza dopravního chování obyvatel Olomouce a Ostravy. *Urbanismus a územní rozvoj*. 2016.

DOBESOVA, Zdena, 2019. The Similarity of European Cities Based on Image Analysis. In: Radek SILHAVY, Petr SILHAVY a Zdenka PROKOPOVA, ed. *Intelligent Systems Applications in Software Engineering*. Cham: Springer International Publishing, s. 341–348. ISBN 978-3-030-30329-7.

DOBESOVA, Zdena, 2020. Experiment in Finding Look-Alike European Cities Using Urban Atlas Data. *ISPRS International Journal of Geo-Information* [online]. 9(6), 406. ISSN 2220-9964. Dostupné z: doi:10.3390/ijgi9060406

DOBEŠOVÁ, Zdena, 2022. *ORANGE, Praktický návod do cvičení předmětu Data Mining*. B.m.: Univerzita Palackého v Olomouci. ISBN 978-80-244-6086-4.

DOBEŠOVÁ, Zdena, Karel MACKŮ a Michal KUČERA, 2022. Výuka geoinformatických předmětů na příkladech dat Evropské Unie. In: *GIS Ostrava 2022. Smart City – vize a realita* [online]. [vid. 2023-03-06]. Dostupné z: doi:10.31490/9788024846071-153

FAYYAD, Usama M., Gregory PIATETSKY-SHAPIRO a Padhraic SMYTH, 1996. From Data Mining to Knowledge Discovery in Databases. *AI Mag.* 17, 37–54.

FOURNIER-VIGER, Philippe, 2023. *SPMF: A Java Open-Source Data Mining Library* [online] [vid. 2023-04-01]. Dostupné z: https://www.philippe-fournier-viger.com/spmf/

FOURNIER-VIGER, Philippe, Jerry Chun-Wei LIN, Bay VO, Tin Truong CHI, Ji ZHANG a Hoai Bac LE, 2017. A survey of itemset mining. *WIREs Data Mining and Knowledge Discovery* [online]. 7(4) [vid. 2023-03-09]. ISSN 1942-4787, 1942-4795. Dostupné z: doi:10.1002/widm.1207

FRAWLEY, William J., Gregory PIATETSKY-SHAPIRO a Christopher J. MATHEUS, 1992. Knowledge Discovery in Databases: An Overview. *AI Mag.* 13, 57–70.

GEURTS, Karolien, Isabelle THOMAS a Geert WETS, 2005. Understanding spatial concentrations of road accidents using frequent item sets. *Accident Analysis & Prevention* [online]. 37(4), 787–799. ISSN 0001-4575. Dostupné z: doi:10.1016/j.aap.2005.03.023

HAN, Jiawei a Micheline KAMBER, 2012. *Data mining: concepts and techniques*. 3rd ed. Burlington, MA: Elsevier. ISBN 978-0-12-381479-1.

JANOUŠEK, Matěj, 2017. *Program pro výpočet plošného indexu v kruhových výsečích* [online]. Olomouc [vid. 2023-03-11]. Univerzita Palackého v Olomouci, Přírodovědecká fakulta. Dostupné z: https://theses.cz/id/kimmu0/?zpet=%2Fvyhledavani%2F%3Fsearch%3Djanou%C5%A1ek%20mat%C4%9Bj%26start%3D1;isshlret=Mat%C4%9Bj%3BJANOU%C5%A0EK%3B

JANOUŠEK, Matěj, 2019. *Porovnání urbánního prostoru pomocí kruhových výsečí* [online]. Olomouc [vid. 2023-02-25]. Univerzita Palackého v Olomouci, Přírodovědecká fakulta. Dostupné z: https://theses.cz/id/w91ot7/

KAGGLE, 2016. Painter by Numbers Competition, 1st Place Winner's Interview: Nejc Ilenič. *Kaggle Blog* [online]. [vid. 2023-04-01]. Dostupné z: https://medium.com/kaggle-blog/painter-by-numbers-competition-1st-place-winners-interview-nejc-ileni%C4%8D-4eaab5e6ce9d

KOPERSKI, Krzysztof a Jiawei HAN, 1995. Discovery of spatial association rules in geographic information databases. In: Max J. EGENHOFER a John R. HERRING, ed. *Advances in Spatial Databases* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, Lecture Notes in Computer Science, s. 47–66 [vid. 2023-03-11]. ISBN 978-3-540-60159-3. Dostupné z: doi:10.1007/3-540-60159-7_4

LOUF, Rémi a Marc BARTHELEMY, 2014. A typology of street patterns [online]. [vid. 2023-03-11]. Dostupné z: doi:10.48550/ARXIV.1410.2094

LUNA, José María, Philippe FOURNIER-VIGER a Sebastián VENTURA, 2019. Frequent itemset mining: A 25 years review. *WIREs Data Mining and Knowledge Discovery* [online]. 9(6) [vid. 2023-03-09]. ISSN 1942-4787, 1942-4795. Dostupné z: doi:10.1002/widm.1329

MIRKO, Gregor, Manuel LÖHNERTZ a Christoph SCHRÖDER, 2018. *Similarities and diversity of European cities. A typology tool to support urban sustainability* [online]. B.m.: European Enviroment Agency. Dostupné z: https://www.eionet.europa.eu/etcs/etc-uls/products/etc-uls-reports/etc-uls-report-03-2018-similarities-and-diversity-of-european-cities-a-typology-tool-to-support-urban-sustainability/@@download/file/etc-uls_report_2018-3-citytypology_final.pdf

PETR, Pavel, 2014a. *Metody Data Miningu část 1*. Vyd. 1. Pardubice: Univerzita Pardubice. ISBN 978-80-7395-872-5.

PETR, Pavel, 2014b. *Metody Data Miningu část 2*. Vyd. 1. Pardubice: Univerzita Pardubice. ISBN 978-80-7395-873-2.

SALMENKIVI, Marko, 2017. Frequent Itemset Discovery. In: Shashi SHEKHAR, Hui XIONG a Xun ZHOU, ed. *Encyclopedia of GIS* [online]. Cham: Springer International Publishing, s. 635–636 [vid. 2023-02-25]. ISBN 978-3-319-17885-1. Dostupné z: doi:10.1007/978-3-319-17885-1_432

ŠARMANOVÁ, J., 2012. *Metody analýzy dat* [online]. Vyd. 1. Ostrava: Vysoká škola báňská - Technická univerzita Ostrava. ISBN 978_80_248_2565-6. Dostupné z: http://www.person.vsb.cz/archivcd/FEI/MAD/MAD.pdf

TRNOVÁ, Lenka, 2020. *Aplikace asociačních pravidel na prostorová data* [online]. Olomouc [vid. 2023-03-07]. Univerzita Palackého v Olomouci, Přírodovědecká fakulta. Dostupné z: https://theses.cz/id/etpi8w/

URBAN ATLAS, 2023. Urban Atlas. *Urban Atlas* [online] [vid. 2023-03-11]. Dostupné z: https://land.copernicus.eu/

URBANČÍK, Filip, 2022. *Podobnost evropských měst a jejich funkčních území* [online]. Olomouc [vid. 2023-02-25]. Univerzita Palackého v Olomouci, Přírodovědecká fakulta. Dostupné z: https://theses.cz/id/fiv1q9/

YUAN, Junsong, Ying WU a Ming YANG, 2007. From frequent itemsets to semantically meaningful visual patterns. In: *KDD07: The 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* [online]. San Jose California USA: ACM, s. 864–873 [vid. 2023-03-09]. ISBN 978-1-59593-609-7. Dostupné z: doi:10.1145/1281192.1281284

ZAKI, M.J., 2000. Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering* [online]. 12(3), 372–390. Dostupné z: doi:10.1109/69.846291

PŘÍLOHY

SEZNAM PŘÍLOH

Volné (elektrické) přílohy

Struktura složky **novak23**, která je odevzdána na digitální uložiště katedry.

Poster

- Novak23.pdf

Text-Prace

- Novak23_text.pdf

Vstupni_Data

- **JupyterNotebook-FI**
  - FI.ipynb
- **MSExcel_soubory**
  - cities100_Janousek.xlsx

Vystupni_Data

- **FREKVENTOVANE_SADY**
  - FS_Czechia_merged
  - FS_Czechia_minsup_5percent
  - FS_EVROPA_minsup_5percent
- **IMG**
  - UA_shluky
  - Europe100_HierarchicalClustering
  - Europe100_HierarchicalClustering_borderline_125
  - Europe100_HierarchicalClustering_borderline_200
- **MSExcel_soubory**
  - PripadovaStudie1_mestaCR.xlsx
  - PripadovaStudie2_mestaEVROPA.xlsx
  - PripadovaStudie2_mestaEVROPA_sumarizacni_matice.xlsx
  - Vyber100_mestaEVROPA.xlsx
- **PROJECT_ArcGISPro**
  - MAIN
    - SearchDistinctLanduse_SpatialJoin.atbx
    - Czechia_UA2018.gdb
    - Europe_UA2018.gdb
- **SCRIPTY**
  - Aggregation
  - FI_ChangeOrder
  - JoinTXT_Files
  - TXT_TO_EXCEL
- **TANSAKCNI_DATA_SOUSEDNOSTI**
  - Czechia_Transactions_Merged
  - Czechia_UA2018_100m
  - Czechia_UA2018_400m
  - Europe_UA2018_100m
  - Europe_UA2018_400m