

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

**POKROČILÁ ANALÝZA A VIZUALIZACE DAT
Z DOTAZNÍKOVÉHO ŠETŘENÍ KATEDRY
GEOINFORMATIKY UP**

Diplomová práce

Bc. Filip FRYČÁK

Vedoucí práce RNDr. Alena VONDRÁKOVÁ, Ph.D., LL.M.

Olomouc 2022
Geoinformatika

ANOTACE

Práce je zaměřena na pokročilou analýzu dat spjatých s Katedrou geoinformatiky Univerzity Palackého v Olomouci a následnou vizualizaci zjištěných poznatků prostřednictvím nástrojů moderního webdesignu. V úvodu práce je pozornost věnována rešerši a kompilaci existujících datových sad. Na základě obsahu užšího výběru datových sad bylo prostřednictvím dotazníkového šetření mezi akademickými pracovníky Katedry geoinformatiky UP stanoveno 12 předpokladů.

V teoretické části práce je stručná rešerše. Obsahem první praktické části práce je pokročilá statistická a prostorová analýza předpokladů, po jejímž provedení byly předpoklady označeny za potvrzené nebo vyvrácené. Jednotlivé analýzy jsou zpracované do podoby přehledových infografik dokumentujících jejich postup, výstupy a vyhodnocení. Součástí jsou i texty diskutující příčiny a výsledky předpokladů.

Druhá část práce je věnována vizualizaci souboru zjištěných poznatků ve webovém prostředí. Vedlejším účelem sestavované prezentace je podpora propagační činnosti Katedry geoinformatiky UP, s orientací na uchazeče o studium. S ohledem na tuto cílovou skupinu je web proveden v atraktivní vizuální podobě a slouží jako bližší představení katedry uchazečům a demonstrace možností uplatnění datové analýzy. Výsledkem práce je mimo jiné seznam doporučení pro další sběr a práci s daty podobného charakteru.

KLÍČOVÁ SLOVA

pokročilá analýza; vizualizace; data; webdesign; Katedra geoinformatiky UP

Počet stran práce: 62

Počet příloh: 11 (z toho 8 vázaných a 3 volné)

ANOTATION

The aim of this thesis is an advanced analysis of data relevant to the Department of Geoinformatics of the Palacký University Olomouc and subsequent visualisation of its findings, utilizing the tools of modern web design. The initial portion of the thesis is comprised of research and compilation of already existing data sets. The result of a narrower selection of these data sets was a survey, carried out among academic staff of the Department of Geoinformatics, which yielded 12 hypotheses.

The theoretical part of the thesis contains a brief research. The first practical part of this work is concerned with advanced statistical and spatial analysis of these hypotheses and as a result of which, these hypotheses were either confirmed or debunked. Each analysis is presented in the form of an infographical overview, showing the process, results, and the evaluation of these results. Paragraphs discussing the causes and results of these hypotheses are also included.

The second part of this work is the visualisation of the findings within a website environment. A secondary goal of this effort is to support the promotion effort of the Department of Geoinformatics, with the target audience being new applicants. In light of this, the website has been built with attractive visuals, and serves to introduce the Department in greater detail to the applicants, while also demonstrating the potential applications of data analysis. A list of recommendations for further collection and analysis of data of similar nature is, among other items, also a result of this work.

KEYWORDS

Advanced analysis; Visualisation; Data; Web design; Department of Geoinformatics UP

Number of pages: 62

Number of appendixes: 11

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, aby jeden výtisk diplomové práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

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

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užití 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 mě požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne 1. 5. 2022

Bc. Filip FRYČÁK

Mé poděkování patří RNDr. Aleně Vondrákové, Ph.D., LL.M za odborné vedení, podněty a čas, který mi při řešení problematiky věnovala. Dále děkuji konzultantům Mgr. Jakubovi Koničkovi za podnětné připomínky v oblasti vizualizace a Mgr. Karlovi Macků, Ph.D. za cenné rady při statistickém vyhodnocení výsledků.

Za poskytnutá data děkuji vedoucímu Studijního oddělení Přírodovědecké fakulty UP Mgr. Jiřímu Mazalovi, pracovníkům Katedry geoinformatiky UP doc. Ing. Zdeně Dobešové, Ph.D., doc. RNDr. Jaroslavovi Burianovi, Ph.D., RNDr. Janu Brusovi, Ph.D. a absolventce studia na Katedře geoinformatiky UP Bc. Marii Šabatové.

V neposlední řadě bych rád poděkoval své rodině, přátelům a přítelkyni Tereze za podporu po celou dobu mého studia.

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2019/2020

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Bc. Filip FRYČÁK**
Osobní číslo: **R190009**
Studijní program: **N1301 Geografie**
Studijní obor: **Geoinformatika**
Téma práce: **Pokročilá analýza a vizualizace dat z dotazníkového šetření Katedry geoinformatiky UP**
Zadávací katedra: **Katedra geoinformatiky**

Zásady pro vypracování

Hlavním cílem diplomové práce je provést pokročilou analýzu dat a následně výsledky prezentovat prostřednictvím nástrojů webové kartografie. První část práce je zaměřena na pokročilou analýzu dat, kdy výchozí datovou sadu tvoří dotazníkové šetření Katedry geoinformatiky UP z roku 2019 (práce M. Šabatové). Student v rámci analýz využije další dostupná data, především kombinaci s daty z Českého statistického úřadu (demografické charakteristiky, dojíždka apod.), z MŠMT (projekt KREDO a další) a z Univerzity Palackého v Olomouci (dotazníkové šetření mezi absolventy, STAG). Součástí teoretické části bude koncepční vymezení, jaké analýzy je možné díky kombinaci uvedených datových zdrojů realizovat. Výsledkem této části práce bude soubor poznatků, a to v podobě získaných údajů, interpretace souvislostí a v podobě mapových výstupů. V druhé části práce se student zaměří na vizualizaci dat a získaných výsledků ve webovém prostředí, kdy výsledkem bude funkční dynamická vizualizace s důrazem na moderní kartografické přístupy. 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:

Šabatová, M. Analýza a vizualizace prostorové diferenciací studentů a absolventů Katedry geoinformatiky UP. Bakalářská práce. Vedoucí práce RNDr. Alena Vondráková, Ph.D., LL.M. Olomouc: Univerzita Palackého v Olomouci, 2019.
Vondráková, A., Voženílek, V., Rychtaříková, J. Atlas vysokého školství v České republice 2001-2014. Olomouc: Univerzita Palackého v Olomouci, 2017. M.A.P.S. (Maps and Atlas Product Series). ISBN 978-80-244-5139-8.
Sojková, A. (Ed.) Výroční zpráva o činnosti Univerzity Palackého v Olomouci za rok 2019. Olomouc: Univerzita Palackého v Olomouci, 2020. ISBN 978-80-244-5685-0.
Voženílek, V. Diplomové práce z geoinformatiky. Olomouc: Univerzita Palackého v Olomouci, 2002. ISBN 80-244-0469-9.
Voženílek, V., Kaňok, J. a kol. Metody tematické kartografie: vizualizace prostorových jevů. Olomouc: Univerzita Palackého v Olomouci, 2011. ISBN 978-80-244-2790-4.

Vedoucí diplomové práce: **RNDr. Alena Vondráková, Ph.D.**
Katedra geoinformatiky

Datum zadání diplomové práce: **7. října 2019**
Termín odevzdání diplomové práce: **6. května 2021**

LS.

doc. RNDr. Martin Kubala, Ph.D.
děkan



prof. RNDr. Vít Voženílek, CSc.
vedoucí katedry

OBSAH

SEZNAM POUŽITÝCH ZKRATEK	9
ÚVOD	10
1 CÍLE PRÁCE	11
2 METODY A POSTUP ZPRACOVÁNÍ.....	12
2.1 Použité metody	12
2.2 Použité technologie	13
2.3 Použitý software.....	14
2.4 Použitá data.....	16
2.5 Postup zpracování.....	17
3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	18
3.1 Stav literatury.....	18
3.2 Katedra geoinformatiky UP.....	20
3.2.1 Popularizace a propagace katedry.....	22
3.3 Data, informace a znalosti.....	22
3.3.1 Datová pyramida.....	23
3.3.2 Záznam a atribut	24
3.3.3 Typy atributů.....	24
3.3.4 Způsoby uložení dat	25
3.4 Datová analýza	27
4 DOSTUPNÉ DATOVÉ ZDROJE	29
4.1 Dotazníkové šetření Vichové.....	29
4.2 Dotazníkové šetření Šabatové.....	29
4.3 Studijní agenda Univerzity Palackého (STAG)	30
4.4 Interní data Katedry geoinformatiky	30
4.5 Externí datové zdroje	31
5 PŘÍPRAVA DAT.....	33
5.1 Výchozí datové sady	33
5.2 Spojení datových modelů	33
5.3 Geokódování	34
5.4 Hrubá filtrace a čištění.....	34
5.5 Transformace.....	34
6 ZÁKLADNÍ STATISTICKÁ ANALÝZA.....	35
6.1 Statistické infografiky	35
6.2 Výsledky statistické analýzy.....	36

7	VYMEZENÍ ANALÝZ	38
7.1	Tematické okruhy	38
7.2	Matice témat	38
7.3	Dotazníkové šetření.....	38
7.4	Stanovení předpokladů	39
8	POKROČILÁ ANALÝZA	40
8.1	Analýza předpokladů	40
8.2	Způsob vyhodnocení	43
8.3	Analytické infografiky.....	44
9	WEBOVÁ VIZUALIZACE	45
9.1	Definování požadavků	45
9.1.1	Účel webu a cílová skupina	45
9.1.2	Technické požadavky	46
9.2	Návrh webu	46
9.2.1	Informační architektura	46
9.2.2	Drátěný model	47
9.2.3	Grafické zpracování.....	47
9.3	Implementace	47
9.3.1	Vývojové prostředí.....	47
9.3.2	Kódování šablony	48
9.3.3	Obsah a optimalizace	48
10	VÝSLEDKY	49
10.1	Dostupné datové zdroje.....	49
10.2	Statistické vyhodnocení	49
10.3	Vyhodnocení předpokladů.....	50
10.3.1	Vyhodnocení předpokladů	51
10.3.2	Analytické infografiky	55
10.4	Webová prezentace.....	57
10.5	Doporučení pro další sběr a práci s daty	58
11	DISKUSE	59
12	ZÁVĚR	62
	POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE	
	PŘÍLOHY	

SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Význam
Adobe CC	Adobe Creative Cloud
Adobe XD	Adobe Experience Design
BEM	Block, Element, Modifier
CEEPUS	Central European Exchange Program for University Studies
CMYK	Cyan, Magenta, Yellow, Key
CSS	Cascading Style Sheets
CSV	Comma-separated Values
DTP	Desktop Publishing
DIKW	Data, Information, Knowledge, Wisdom
ERASMUS	European Action Scheme for the Mobility of University Students
ETL	Extraction, Transformation, Loading
HTML	HyperText Markup Language
IS/STAG	Informační systém studijní agendy vysoké školy
IZO	Identifikační znak organizace
JSON	JavaScript Object Notation
KGI	Katedra geoinformatiky Univerzity Palackého v Olomouci
KREDO	Kvalita, relevance, efektivita, diverzifikace a otevřenost
MIT	Massachusetts Institute of Technology
MŠMT	Ministerstvo školství, mládeže a tělovýchovy
ODS	Open Document Spreadsheets
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
SASS	Syntactically Awesome Style Sheets
SCSS	Sassy Cascading Style Sheets
SEO	Search Engine Optimization
SIMS	Sdružené informace matrik studentů
SQL	Structured Query Language
UP	Univerzita Palackého v Olomouci
XSL	Excel Spreadsheets
YALE	Yet Another Learning Environment (<i>RapidMiner</i>)

ÚVOD

V dnešním světě data vznikají závratnou rychlostí. Ať už se hovoří o datech manuálně nebo automaticky sbíraných, s jejich rostoucími objemy se přímo nabízí otázka, jaký je smysl a cíl jejich ukládání. Data jsou pouze surovými údaji reprezentujícími vlastnosti objektů a jevů (Ackoff, 1989). Samy o sobě nejsou nositelem žádného významu a nejsou nikterak užitečná. Je potřeba v nich bádát, zkoumat je a nalézat v nich vzory a souvislosti. Teprve tak mohou mít nashromážděná data skutečný přínos, být užitečná a přinášet cenné odpovědi. Z dat vyplývající znalosti mohou následně sloužit například pro podporu rozhodování nebo strategické plánování. Anglicky se takový proces označuje jako *data driven decision making*. Nástrojem pro hledání nových souvislostí a vzorů v datech je právě datová analýza.

Shromažďování dat již dávno není doménou pouze akademickou. Informační vědy stále častěji pronikají i do světa komerce. Potenciální přínosy rozhodování na základě dat si uvědomuje čím dál více společností a firem, marketingový význam si uvědomují i vysokoškolské instituce. Nová data rychle vznikají i v rámci studijních agend, dotazníkových šetření nebo akademických prací. Katedra geoinformatiky Univerzity Palackého v Olomouci díky své prominentní akademické i mimoakademické činnosti disponuje řadou datových sad, v nichž se, z hlediska potenciálního uplatnění, nepochybně ukrývá množství cenných odpovědí, ale i skutečností vyvolávajících další otázky. Možnosti využití znalostí obsažených v datech jsou téměř neomezené. Oblasti uplatnění sahají od strategického rozhodování po oblast popularizace a propagace.

Tato diplomová práce přímo navazuje na bakalářskou práci *Analýza a vizualizace prostorové diference studentů a absolventů Katedry geoinformatiky UP* (Šabatová, 2019). Šabatová se v rámci své práce zabývala sběrem informací o uchazečích, studentech a absolventech KGI a následnou prezentací získaných údajů ve formě propagační brožury. Cílem této práce je zkoumat další možnosti analýzy existujících dat a hledat nové způsoby přiblížení získaných výsledků široké veřejnosti.

1 CÍLE PRÁCE

Hlavním cílem diplomové práce je provést pokročilou analýzu dat a následně výsledky prezentovat prostřednictvím nástrojů webové kartografie.

První část práce je zaměřena na pokročilou analýzu dat, kdy výchozí datovou sadu tvoří dotazníkové šetření Katedry geoinformatiky UP z roku 2019, které bylo realizováno v rámci bakalářské práce Marie Šabatové (2019). Je provedena rešerše dalších existujících datových sad spjatých s Katedrou geoinformatiky UP a dat potenciálně přínosných v nadcházejících analýzách. Nejsou opomenuty zejména data ze Studijní agendy Univerzity Palackého v Olomouci, data z Českého statistického úřadu (demografické charakteristiky, dojíždka apod.) a data z Ministerstva školství, mládeže a tělovýchovy (např. projekt KREDO). Výstupem této fáze je seznam všech dostupných datových sad, z nichž jsou dále selektovány sady pro analýzy.

Součástí teoretické části práce je koncepční vymezení, jaké analýzy je možné díky kombinaci uvedených datových zdrojů realizovat. Jsou definována klíčová témata, jejichž analýza přináší ověření všeobecně míněných předpokladů a objevení nových překvapivých souvislostí. Důraz je kladen, mimo jiné, na uplatnitelnost zjištění v praxi, např. v oblasti popularizace a propagace katedry. Výsledkem této části práce je soubor poznatků, a to v podobě získaných údajů, interpretace souvislostí a v podobě mapových výstupů.

Druhá část práce je zaměřena na vizualizaci dat a získaných výsledků ve webovém prostředí. Autor zde prezentuje zjištěné výsledky, a to s ohledem na dále stanovenou cílovou skupinu a její zájmy. Cílová skupina je stanovena na základě charakteru zjištěných poznatků. Web má podobu funkční, interaktivní a dynamické vizualizace s důrazem na moderní kartografické přístupy, a splňuje současné standardy webové tvorby (responzivní, multiplatformní, uživatelsky přívětivý apod.).

Vedlejším cílem práce je sumarizovat poznatky nabyté při zpracovávání existujících dat a vytvořit seznam doporučení pro jejich další sběr a práci s nimi. Sestavená doporučení by měla umožnit dalším autorům na odvedenou práci snadno navázat a aplikovat poznatky do vlastní vědecké činnosti.

Práce si klade za cíl provést rešerši dostupných datových sad, sady popsat a selektovat z nich potenciálně nejprínosnější, provést jejich analýzu a odprezentovat zjištění formou atraktivní webové vizualizace. Ta by mohla najít uplatnění mimo jiné např. v oblasti popularizace katedry, či jako demonstrace dosažitelných poznatků, a případně jako motivace pro uplatnění analytických metod ve větším měřítku (např. fakultním, univerzitním).

2 METODY A POSTUP ZPRACOVÁNÍ

Diplomová práce prostupuje výčtem vědních disciplín. V každé z nich byla využita řada oborově známých a standardně používaných metod, technologií a softwarů. Níže je uveden jejich přehled a stručný popis. O metodách a způsobu jejich aplikace se lze blíže dozvědět v dalších kapitolách práce, nebo v citované literatuře.

2.1 Použité metody

Online dotazník

Pro výběr témat vhodných k analýze byl v prostředí *Google Sheets* sestaven jednoduchý online dotazník, který byl rozeslán vybraným akademickým pracovníkům Katedry geoinformatiky UP pro zjištění jejich expertního názoru. Dotazník měl podobu křížové matice dostupných témat, přičemž respondenti označovali kombinace považované za potenciálně přínosné či zajímavé. Kromě vyznačení kombinace byli vyzváni i k přiložení stručného komentáře zdůvodňujícího jejich výběr. Přestože je nástroj *Google Sheets* tabulkový procesor a není primárně určen pro tvorbu dotazníku, byl vybrán pro svou jednoduchost a možnost cloudového sdílení. Protože se počet dotazovaných respondentů pohyboval v řádech jednotek, vyplněné dotazníky nebylo nutné sbírat s využitím sofistikovanějšího nástroje.

Delfská metoda

Témata vhodná k analýze byla stanovena pomocí metody získávání expertních výpovědí, tzv. delfskou metodou. Tu lze charakterizovat jako písemné dotazování (na základě formalizovaných dotazníků) založené na postupném zjišťování prognóz (názorů) expertů (Šulc, 1987). Spolu s respondenty byly kombinace ohodnoceny i autorem práce. Každé kombinaci témat bylo jednoduchou bodovací metodou podle počtu hlasů přiřazeno hodnocení. Hlasy všech respondentů i hlas autorův měly rovnocennou váhu. Kombinace témat s nejvyšším počtem bodů byly vybrány k analýze.

Metody tematické kartografie

Při analýze dat a tvorbě mapových výstupů byla využita celá řada konvenčních metod tematické kartografie. Nejčastěji byly při vizualizacích voleny metody bodové a areálové. Použité metody byly vybrány s ohledem na charakter analyzovaných dat, zejména přepočítání kvantitativního jevu na jednotku plochy, a znázornění kvantitativních jevů v absolutních hodnotách. Cílem této práce není dopodrobna metody popsat, proto je zde uveden pouze jejich výčet. Použité metody blíže popisují Voženilek, Kaňok a kol. v publikaci *Metody tematické kartografie* (2011). Výčet je následující:

- metody bodových, liniových a plošných znaků,
- metoda kartogramu a kartodiagramu,
- metoda gridu.

Mood board

Při tvorbě grafického návrhu webové prezentace byl vytvořen tzv. *mood board*, neboli koláž grafických prvků, barev a dalších designových inspirací vytvářejících celkový dojem, jenž by měl plynout z vytvořeného produktu (Trulock, 2008). Metoda umožňuje znázornit emoce a hodnoty, které lze slovy obtížně vyjádřit. *Mood board* nachází uplatnění v mnoha designérských oborech, v neposlední řadě v oboru grafického návrhu webových stránek a dalších digitálních produktů.

Block, Element, Modifier (BEM)

Block, Element, Modifier (zkráceně BEM) je metoda organizace HTML a CSS kódu. Jedná se o přístup webového vývoje kaskádových stylů založený na rozdělení uživatelského rozhraní do nezávislých bloků (Yandex, 2022). Metoda zavádí konvence v pojmenovávání komponent, umožňující rychlejší a snadnější vývoj webových stránek, s maximálním využitím existujících kaskádových stylů obsažených v projektu. Princip metody spočívá v rozdělení struktury rozhraní do nezávislých bloků (*block*), obsahujících vlastní prvky (*element*). Bloky se mohou v projektu vyskytovat v různých podobách a stavech, které jsou definovány jako modifikace původního bloku (*modifier*).

Architektura 7-1

Architektura 7-1 (nebo také vzor 7-1) je oficiálně nezdokumentovaným, avšak *de facto* standardním a konvenčním přístupem k organizaci a strukturování kaskádových stylů uvnitř webového projektu. Metoda je založena na principu rozdělení CSS komponent do sedmi adresářů, lišících se úrovní abstrakce obsažených komponent. Adresáře mohou být podle potřeby vynechány, nebo upraveny vzhledem k požadavkům konkrétnímu projektu. Komponenty jsou poté načítány do jednoho hlavního souboru. Takové členění umožňuje naplno využívat dědičnosti CSS komponent a současně udržovat přehlednost v psaném kódu. Výsledkem je minimální a snadno škálovatelná CSS architektura. Metoda 7-1 je blíže popsána v komunitní dokumentaci technologie SASS, SASS Guidelines (2022), a v mnohých neoborných publikacích (Timothy Robards, Denis Buchenko a další).

2.2 Použité technologie

HTML, CSS a JavaScript

Značkovací jazyk *HyperText Markup Language* (HTML), jazyk pro definici grafického způsobu zobrazení *Cascading Style Sheets* (CSS) a programovací jazyk *JavaScript* jsou trojice technologií, bez kterých si současný webový vývoj jen stěží představit. Jedná se o základní stavební kameny téměř každého webu. Při realizaci rozsáhlejších projektů se ale psaní kódu těmito technologiemi v jejich nativní podobě stává obtížným a časově náročným. Proto je vhodné využít další nadstavbové technologie, které umožňují zmíněným neduhům předcházet, a v konečném důsledku vytvořit kvalitnější web.

Jedním z hlavních problémů při psaní kódu může být jeho nekontrolovaný nárůst. Soubor o stovkách či tisících řádků se nejen těžko spravuje, ale i umocňuje pravděpodobnost zanesení chyby do kódu vlivem lidského faktoru. Proto je vhodné zdrojový kód členit do atomických (nebo alespoň menších) částí, které jsou následně kompilovány do výsledného kódu určeného k produkci.

Nunjucks

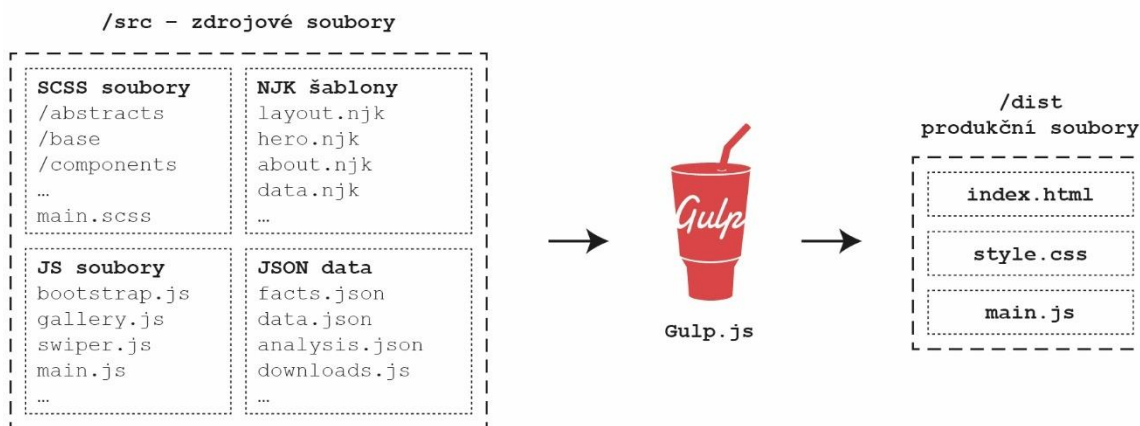
Nunjucks je HTML šablonovací jazyk, umožňující komponentizovat zdrojový HTML kód. Kromě samotného rozdělení kódu do samostatných souborů umožňuje i použití cyklů, tvorbu proměnných, zahrnutí logických podmínek a další (Mozilla GitHub, 2022). Díky těmto vlastnostem lze mimo jiné kompletně oddělit strukturní a datovou složku, a data do HTML dokumentu načítat z externího souboru, určeného přímo k jejich uchovávání (např. JSON). Výhodou jazyku *Nunjucks* je podobnost syntaxe s jazykem HTML.

Syntactically Awesome Style Sheets (SASS)

SASS je rozšířením jazyka CSS o celou řadu užitečných funkcí, které nativní verze CSS neposkytuje. Jmenovitě se jedná např. o zjednodušení zápisu proměnných, hníždění (angl. *nesting*), komponentizaci kódu do samostatných souborů (angl. *partials*), tvorbu funkčních bloků (angl. *mixins*) (SASS tým a přispěvatelé, 2022). Všechny tyto funkce vedou k rychlejšímu, úspornějšímu a lépe škálovatelnému zápisu zdrojového kódu. Výhodou je zpětná kompatibilita s jazykem CSS. V oborové praxi i této práci je používána jeho modifikace SCSS (*Sassy CSS*), která se syntaxí blíží nativní podobě CSS.

Gulp.js

Gulp je nástrojem pro automatizaci úloh, které by webový vývojář musel provádět manuálně, často opakovaně. V zásadě se jedná o kompilaci zdrojových souborů do výsledné distribuční podoby, během které dochází k předem definovaným úlohám. Typickými úlohami jsou převod kódu z kompilovaného jazyka do jazyka nativního, optimalizace kódu pro různé prohlížeče, minifikace apod. (obr. 1). Gulp sám o sobě je pouze platformou pro automatizaci, jednotlivé úlohy jsou prováděny prostřednictvím importovaných modulů. Nástroj byl použit právě pro kompilaci zdrojových souborů dříve popsanych technologií do nativních formátů HTML, CSS a JavaScript.



Obr. 1 Schéma vývoje webové prezentace.

2.3 Použitý software

RapidMiner Studio (verze 9.10.001, Educational Edition)

RapidMiner je platforma produktů poskytujících nástroje pro přípravu dat, manipulaci s daty, strojové učení, data mining, prediktivní analytiku a spoustu dalších metod datových věd. Je vyvíjena stejnojmennou firmou. Počátky platformy sahají do roku 2001, kdy byl na Technické univerzitě v Dortmundu zahájen její vývoj, tehdy pod názvem *YALE* (RapidMiner, 2022). V současnosti je platforma komerčním projektem, orientovaným kromě samotného výzkumu i na komerční sféru. Vlajkovým produktem společnosti je software *RapidMiner Studio*. V tomto robustním softwaru lze ve vizuálním editoru nad daty vytvářet libovolné procesy a modely, od jednoduchých úprav atributů po sofistikované metody datových věd. Důležitým aspektem je přitom zachování výchozí datové sady, ze které model vychází, a do které není třeba destruktivně zasahovat. Uživatel v editoru definuje posloupnost jednotlivých operací do podoby modelu. Vytvořený model je následně i samotnou dokumentací celého procesu. V *RapidMiner Studio* probíhala téměř veškerá příprava.

ArcGIS 10.4.1 for Desktop

Jevy s prostorovým aspektem byly zpracovány a analyzovány v *ArcGIS for Desktop*. Jedná se o komplexní geografický informační systém poskytující celou řadu nástrojů pro práci s prostorovými informacemi. V práci byl použit pro prvotní převod dat z textových formátů do vektorových formátů geografické polohy a pro analytické operace nad daty. Byly v něm vytvořeny podklady, vstupující do dále použitých grafických softwarů.

Adobe Creative Cloud

Adobe Creative Cloud je balíček více než 20 kreativních aplikací a služeb určených pro grafický design, návrh webů a uživatelských rozhraní, tvorbu videí, úpravu fotografií a další. Obsahuje také základní zdroje, jako jsou palety barev nebo rodiny písem (Adobe, 2022). Umožňuje i sdílení grafických podkladů nejen mezi uživateli, ale i napříč vlastními aplikacemi. Tímto způsobem byla v práci spravována knihovna obsahující barvy, písma, styly odstavců, ikony apod. Balíček aplikací byl zakoupen ve studentském plánu formou měsíčního předplatného.

Adobe Illustrator CC 2022 (verze 26.1)

Výstupy analýz byly vizuálně zpracovány v softwaru *Adobe Illustrator*. Tento vektorový grafický editor poskytuje pokročilé nástroje pro tvorbu tiskové i digitální grafiky, využitelné mimo jiné v mapové tvorbě. V *Adobe Illustratoru* byly zpracovány veškeré grafy, infografiky i mapové výstupy, a to buď kompletně, nebo *Illustrator* sloužil jako postprocesor prvotních exportů ze softwarů použitých při analýzách.

Adobe InDesign CC 2022 (verze 17.1)

Tvorba přehledových infografik dat i prezentačních infografik výsledků analýz probíhala v DTP softwaru *Adobe InDesign*. Oproti *Illustratoru* je *InDesign* více zaměřen na tvorbu rozsáhlých dokumentů. Výhodami jsou např. svižnost softwaru při práci s datově objemnými soubory, šablonování formou tzv. rodičovských stránek, nebo pokročilé funkce správy barev, typografie apod.

Adobe Experience Design (verze 49.0.12)

Adobe Experience Design (známý také jako Adobe XD) je vektorově založený grafický editor pro tvorbu uživatelských rozhraní. Kromě designování samotné grafiky umožňuje také výrobu interaktivních prototypů. V práci byl použit pro grafický návrh rozhraní prezentační webové stránky.

Další použitý software

Prvotní seznámení s daty a jejich hrubé předzpracování bylo provedeno v tabulkovém procesoru *Microsoft Excel* (kancelářský balík *Microsoft 365 Apps*, školní licence). Tvorba a sběr online dotazníků zaslaných pracovníkům katedry proběhla v cloudovém tabulkovém procesoru *Google Sheets* (kancelářský balík *Google Docs*, neplacený plán). Geokódování souřadnic bylo rovněž provedeno v *Google Sheets* prostřednictvím zásuvného modulu (skriptu) *Awesome Table* (plán Free). Kódování a vývoj webové prezentace probíhal v integrovaném vývojovém prostředí *Visual Studio Code* (verze 1.65.2, User Setup). Pro správu JavaScript knihoven použitých při vývoji byl použit správce balíčků *Node.js Package Manager* (verze 8.5.5), instalován byl prostřednictvím systému *Node.js* (verze 16.14.2). Do vývoje webu byla zahrnuta řada knihoven, jejich kompletní výčet včetně verzí je uveden v příloze 8. *Přehled vývojových JavaScript balíčků*. Verzování zdrojového kódu bylo zajištěno technologií *Git* (verze 2.35.1) s hostováním na webové službě *Bitbucket* (plán Free).

2.4 Použitá data

Dotazníkové šetření Šabatové (2019)

Výchozí datovou sadou bylo dotazníkové šetření Marie Šabatové z roku 2019, vytvořené v rámci bakalářské práce *Analýza a vizualizace prostorové diferenciace studentů a absolventů Katedry geoinformatiky UP*. Šabatová prostřednictvím online dotazníku posbírala údaje o uchazečích, studentech a absolventech KGI. Data byla převzata jako výchozí export z nástroje *JotForm*. Vstupem ke zpracování byly tři samostatné soubory uložené ve formátu CSV.

Studijní agenda UP

Stěžejními byla i data o uchazečích a studentech ze Studijní agendy UP, poskytnutá Studijním oddělením Přírodovědecké fakulty Univerzity Palackého v Olomouci, zastoupeným Mgr. Jiřím Mazalem. Data o uchazečích byla dostupná za akademické roky 2003/04–2019/20, data o studentech za akademické roky 2000/01–2019/20. Data byla po jednotlivých akademických letech uložena jako souborové formáty XLS.

Interní data Katedry geoinformatiky UP

Do analýz byly zahrnuty dvě datové sady interně vedené Katedrou geoinformatiky UP. První sadou byla data o uchazečích v akademických letech 2016/17–2019/20, vedená za účelem akademického výzkumu, a poskytnuta doc. Ing. Zdenou Dobešovou, Ph.D. Druhou sadou jsou záznamy o účastnících zahraničních mobilit, dostupné za roky 2004/05–2021/22, a poskytnuta doc. RNDr. Jaroslavem Burianem, Ph.D.

ArcČR® 500 (verze 3.3)

Jako podkladová data pro analýzy byla použita digitální vektorová databáze České republiky *ArcČR® 500* verze 3.3 od společnosti *ARCDATA PRAHA, s. r. o.* Přestože v okamžik zpracování již byla dostupná aktuálnější verze 4.0, s ohledem na časové určení analyzovaných dat byla zvolena starší verze 3.3. Podkladová data zahrnutá do analýz jsou platná k 1. 1. 2015.

Další použitá data

Doplňujícími daty byl *Rejstřík škol a školských zařízení* poskytovaný Ministerstvem školství, mládeže a tělovýchovy, ve stavu k roku 2021, a některé demografické charakteristiky obyvatelstva z Českého statistického úřadu. Všechny datové sady jsou blíže popsány v kapitole 4. *Dostupné datové zdroje*.

2.5 Postup zpracování

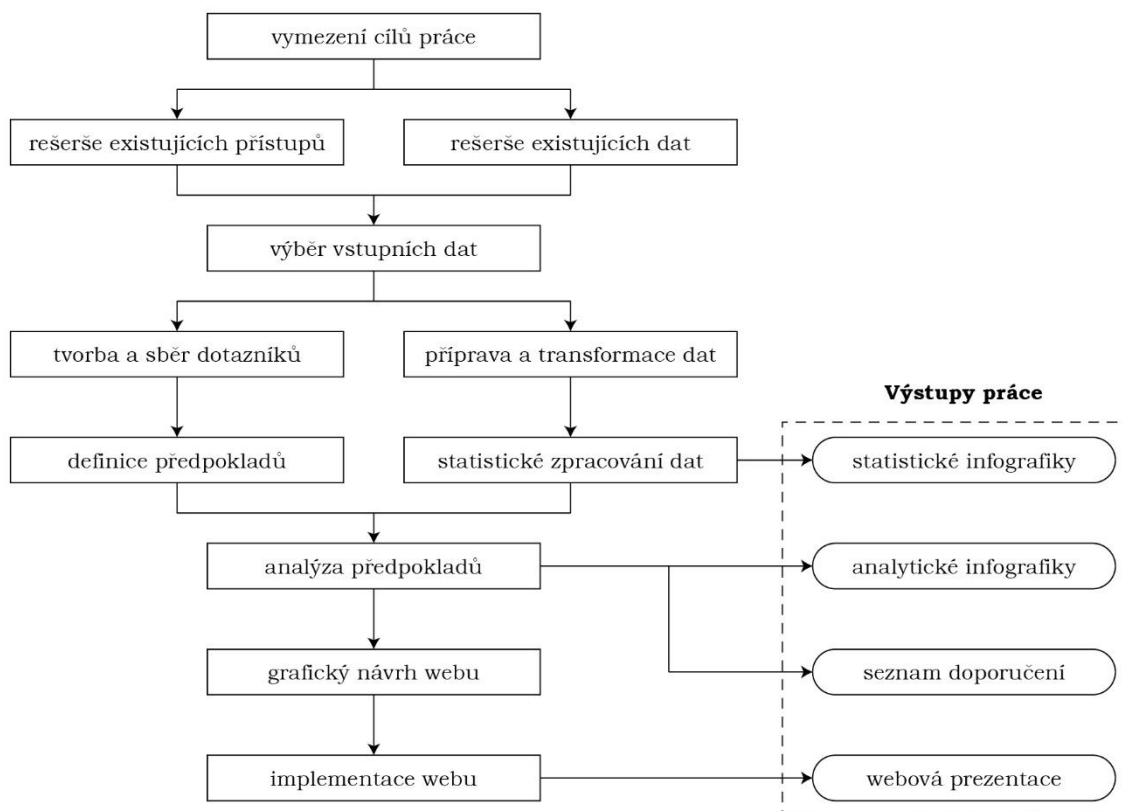
Práce postupovala celým cyklem získání informace od kompilace dostupných datových sad přes jejich analýzu až po jejich vyhodnocení, interpretaci a grafickou prezentaci. Po vymezení cílů byl postup rozpracován do čtyř na sebe navazujících fází.

V první fázi byla provedena rešerše existujících datových sad potenciálně přínosných v nadcházejících analýzách, a rešerše existujících přístupů k analýze podobného typu dat. Následně byly všechny datové sady zkompilovány. Z takto vytvořeného souboru sad byly selektovány sady vhodné k zahrnutí do analýz.

Ve druhé fázi proběhlo bližší seznámení se s daty. Na základě průzkumu dat byly definovány okruhy témat a vytvořena jejich křížová matice. Matice byla formou dotazníku zaslána expertním respondentům. Bylo stanoveno celkem 12 předpokladů k analýze. Ke konci druhé fáze byly vypracovány přehledové statistické infografiky datových sad.

Třetí fází byla samotná analýza stanovených předpokladů a jejich vyhodnocení. Výstupem této fáze jsou výsledkové analytické infografiky podávající přehled o postupu, mezivýsledcích, vyhodnocení a diskusi ke každému z předpokladů. Současně byl vypracován přehled doporučení pro další sběr a práci s těmito daty.

Cílem poslední fáze bylo sestavení atraktivní webové vizualizace prezentující dosažené výsledky. Nejprve byly definovány požadavky na web a proveden výběr technologií pro jeho vývoj. Následně proběhla samotná implementace.



Obr. 2 Schéma postupu práce.

3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

3.1 Stav literatury

Tato diplomová práce přímo navazuje na práci *Analýza a vizualizace prostorové diferenciacie studentů a absolventů Katedry geoinformatiky UP* (Šabatová, 2019). Šabatová se ve své práci zabývala analýzou dat z existujících datových zdrojů, zejména dat ze Studijní agendy Univerzity Palackého v Olomouci, i nově vytvořených dat vyplývajících z provedeného dotazníkového šetření mezi uchazeči, studenty a absolventy Katedry geoinformatiky UP. Prostřednictvím online dotazníkového šetření Šabatová získala množství dat různých tematických kategorií, např. základní demografické údaje, způsob bydlení, mobility, hodnocení studia, zaměstnání. Získaná data byla zpracována, nad zpracovanými daty provedeny četné prostorové i neprostorové analýzy a vytvořeny kartografické a infografické vizualizace. Výsledky a závěry své práce Šabatová přehledně zpracovala do podoby propagační infograficko-mapové brožury pod názvem *Katedra geoinformatiky ...do hlubin duše*.

Sběrem dat a analýzou prostorové diferenciacie studentů Katedry geoinformatiky UP se též zabývala Vichová (2007) v práci *Prostorové analýzy uchazečů, studentů a absolventů geoinformatiky na UP Olomouc*. Podobně jako Šabatová čerpala data ze Studijní agendy Univerzity Palackého v Olomouci a vlastního dotazníkového šetření mezi uchazeči, studenty a absolventy bakalářského a magisterského studia. Sestavila celkem devět dotazníků. Autorka uvádí, že s cílem neovlivňovat respondenta bylo do dotazníků záměrně začleněno množství otevřených otázek vedoucích respondenta k hlubšímu zamyšlení, a tedy předpokladu vyšší vypovídací hodnoty dotazníků. Získaná data Vichová zpracovala a uložila do relační databáze. Nad databází provedla množství prostorových i neprostorových analýz. Analýzy byly rozděleny podle varianty dotazníku do dvou částí, tj. podle dotazníků předkládaných v letech 1997–2006 a dotazníku předloženého v roce 2006. Hlavním cílem prostorových analýz bylo sledovat, kam směřují absolventi geoinformatiky. Tento cíl ale z důvodu nízké návratnosti dotazníků nebyl naplněn. Závěrem pak Vichová stanovila tvrzení, zejména pak o časovém vývoji charakteristik uchazečů, studentů a absolventů.

Základní prostorovou diferenciací a vnitřní strukturou uchazečů o studium na Přírodovědeckou fakultu Univerzity Palackého v Olomouci (dále jen PŘF UP) se zabývala Hakenová (2009) v práci *Prostorová a oborová diferenciacie uchazečů o studium na PŘF UP v Olomouci*. Data ze Studijní agendy Univerzity Palackého v Olomouci Hakenová analyzovala především z hlediska demografických údajů, úspěšnosti u přijímacích zkoušek a oborové struktury. Napříč celou prací porovnávala data z akademických let 2007/2008 a 2008/2009 a studuje jejich vývoj. Výsledky Hakenová vizualizovala prostřednictvím grafů a několika kartografických výstupů.

Pokorný (2010) se ve své práci *Zpracování výsledků dotazníkového šetření absolventů Přírodovědecké fakulty Univerzity Palackého v Olomouci* zabýval uplatněním absolventů PŘF UP na trhu práce. V úvodu stanovil základní výzkumné otázky. Následně formou dotazníkového šetření oslovil absolventy PŘF UP, odpovědi otestoval prostřednictvím statistických hypotéz a vyhodnotil jejich platnost. Návratnost dotazníků Pokorný uvedl 24 % a vzhledem k délce trvání šetření (1 měsíc) jej považuje za velmi dobrý výsledek.

Analogický vzor postupu lze pozorovat v práci Petráše (2011) *Prostorová diferenciacie studentů Univerzity Palackého v Olomouci*. Sestavený dotazník zamýšlel předložit studentům u zápisu ke studiu, z technických důvodů však k faktickému předložení došlo až v průběhu zimního semestru 2010. Po důkladném vymezení postavení Univerzity Palackého a PřF UP v rámci vysokého školství v České republice statisticky neprostorově zhodnotil strukturu jejich studentů. Následně výsledky vizualizoval i prostřednictvím kartografických výstupů, zejména metodou kartogramů, a sice v poměru počtu studentů jednotlivých fakult na obyvatele.

Problematikou prostorových analýz studentů se zabýval i Pučan (2010) v práci *Prostorové analýzy studentů a absolventů oboru Geoinformatika na VŠB-TU Ostrava*. Mezi studenty a absolventy provedl dotazníkové šetření, ze kterého následně analyzoval několik tematických oborů (profil studenta, hodnocení studia, motivace ke studiu, představa o budoucím zaměstnání, uplatnění absolventů atd.). V práci se více věnoval diferenciaci prostorové, a tedy na tuto část kladl větší důraz, než je v pracích podobné tematiky obvyklé. Cílem práce bylo získat náměty na změny a zkvalitnění výukového procesu oboru Geoinformatika na Vysoké škole báňské – Technické univerzitě Ostrava.

Další práci zabývající se analýzou studentů z prostorového hlediska je *Prostorové chování studentů institutu geoinformatiky* (Linhartová, 2016) provedená rovněž na VŠB-TU Ostrava. Linhartová se zabývala analýzou z hlediska krátkodobého pohybu studentů (v rámci měsíců), v práci ale uplatnila metody explorační a prostorové explorační analýzy, které lze aplikovat i na data nižších časových rozlišení a delších časových úseků. V práci se potýkala s nízkou participací a vysokou experimentální mortalitou. Z těchto důvodů Linhartová uvádí, že práci lze považovat spíše za návod či metodický postup při zpracování časoprostorových dat.

Tématu úzce souvisejícím s prostorovou diferenciací studentů a porozumění jejich chování se věnovala Schneiderová (2017) v práci *Dojížděka studentů ekonomických fakult za studiem v Brně*. Data čerpala z databáze SIMS (Sdružené informace matrik studentů). Pro jednotlivé ekonomické fakulty v Brně vymezila regiony dojížděky. Současně porovnávala vývoj mezi lety 2000 a 2016. Výsledky vizualizovala prostřednictvím kartografických výstupů. Závěrem statisticky sumarizovala zjištěné poznatky a popsala příčiny změn ve sledovaném období.

Dojížděkou studentů do vysokých škol v Brně se rovněž zabývala Altmannová (2014). V práci se soustředí především na zahraniční studenty. Využívá zejména dat o dojížděce z Českého statistického úřadu a dat z Ministerstva školství, mládeže a tělovýchovy. Dále realizovala dotazníkové šetření mezi studenty ze Slovenska, na jehož základě zkoumá příčiny a motivaci dojezdu těchto studentů do vysokých škol v Brně. Jako za jednu z hlavních příčin Altmannová považuje nedostatečnou nabídku tamních vysokých škol. Tvrzení dokládá sníženou poptávkou po studiu z takových slovenských regionů, kde existuje oborově srovnatelná nabídka vysokých škol.

Obecněji se prostorovou analýzou školských dat zabýval Rykr (2013) v práci *Prostorové analýzy a vizualizace školských dat Olomouckého kraje*. Data byla poskytnuta z databáze Školské matriky Olomouckého kraje za rok 2010/2011. Nad daty byly provedeny analýzy a následně vizualizace. Dále se pak Rykr zabýval spíše návrhem a tvorbou *Atlasu škol Olomouckého kraje* s důrazem na aplikaci principů systematické kartografie.

Mimo uvedené kvalifikační práce se řešené problematice věnuje v národním měřítku Ministerstvo školství, mládeže a tělovýchovy, které analyzuje prostorovou diferenciaci žáků a studentů, dojížděku do škol apod. Rozsáhlé analýzy proběhly například v rámci projektu IPN KREDO (Individuální projekty národní pro oblast terciárního vzdělávání, výzkumu, vývoje a inovací – *Kvalita, relevance, efektivita, diverzifikace a otevřenost vysokého školství v ČR. Strategie vysokého školství do roku 2020 v návaznosti na Evropu 2020 a nový zákon o vysokých školách*). Prostorové analýzy a vizualizace byly vytvářeny i na Katedře geoinformatiky UP, a to k problematice vysokého školství v ČR. Výsledky byly publikovány především v atlase *Atlas vysokého školství v České republice 2001–2014* (Vondráková a kol., 2022). Byla zde publikována i případová studie na Univerzitu Palackého v Olomouci, přičemž data byla agregována za jednotlivé fakulty.

V zahraničí se analýzy nad různorodými daty o studentech realizují také, přičemž výsledky jsou tematicky zaměřené na různé oblasti, jako jsou faktory úspěšnosti studentů, jejich demografickou variabilitu, dojížděku do škol, genderové rozdíly apod. (Nordvik a Amponsah, 1998; Roessingh, 2008; Pike a kol., 2011; Hendrickson a kol., 2011; Kamruzzaman a kol., 2011). Analýzou prostorové charakteristiky středoškolských studentů se zabýval například Heo a kol. (2017). Vnitrostátní i mezinárodní mobilitou studentů v Itálii se ve své práci zabýval Columbu a kol. (2021). Otázkami, jakou roli ve výběru školy hrají nejen prostorové charakteristiky, se zabývali Lubienksi a Dougherty (2009). Možnosti uplatnění GIS a prostorového modelování v plánování školského systému popsali Clark a Langley (1996).

Jedním z možných scénářů nadcházející analytické části této práce je stanovení předpokladů a jejich následné ověření. V takovém případě by nemělo být opomenuto exaktní statistické vyhodnocení a nasazení odpovídajících metod. Statistickým vyhodnocením prostorových závislostí a predikcí kategorických závislých proměnných prostřednictvím logistické regrese se zabývali například Řehák a Řeháková v práci *Logitové modely: analýza vlivu exogenních faktorů u kategorizovaných dat* (1992).

Na základě provedené rešerše současného poznání a prací s totožnou nebo příbuznou tematikou lze identifikovat několik fází procesu analýzy dat o studentech. V první řadě je nezbytné disponovat odpovídajícími daty. Taková data jsou často tvořena formou dotazníkového šetření. Druhou fází procesu je seznámení se s daty prostřednictvím základní statistické analýzy. V pokročilejších pracích jsou data analyzována prostorově, či doplněna o externí data ze studijních agend univerzit, studijních matrik, statistických úřadů apod. Typickými výstupy jsou diagramy znázorňující demografickou či oborovou diferenciaci studentů či vymezení regionů dojížděky zpracovány do podoby kartografické vizualizace. Výjimečně je v závěru prací uvedena úvaha nad příčinou zjištěné diferenciaci a možnostmi aplikace poznatků do praxe.

3.2 Katedra geoinformatiky UP

Katedra geoinformatiky Univerzity Palackého v Olomouci je jedním z českých univerzitních akademických pracovišť zabývajících se geoinformačními systémy, dálkovým průzkumem Země, kartografií, modelováním prostorových jevů a oborů příbuzných. V rámci organizační struktury Univerzity Palackého v Olomouci spadá katedra pod Přírodovědeckou fakultu. Společně s Katedrou geografie, Katedrou geologie a Katedrou rozvojových a environmentálních studií tvoří Katedra geoinformatiky obor Vědy o Zemi.

Počátky výuky geoinformatiky na Univerzitě Palackého v Olomouci sahají do roku 1992, kdy se poprvé na Univerzitě Palackého, tehdy v rámci katedry geografie, objevilo první programové vybavení GIS (Voženílek, 2011). Do výuky byly postupně zařazovány nové geoinformačně orientované předměty. Rostoucí atraktivita oboru a akademické úspěchy otevřely olomoucké geoinformatice cestu k předložení žádosti o akreditaci pro bakalářské studium. Akreditace byla úspěšně získána v roce 1997 pod názvem Geografie a geoinformatika. Samostatná Katedra geoinformatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci byla rozhodnutím akademického senátu založena k 1. září 2001, se sídlem na adrese tř. Svobody 26, Olomouc. Katedra postupně získala akreditaci pro navazující magisterské studium a rigorózní zkoušku (2002), doktorské studium (2007) a habilitační a profesorské řízení (2016). V roce 2013 došlo k přesídlení katedry na novou adresu 17. listopadu 50, Olomouc.

V současnosti katedra zajišťuje výuku geografické a tematické kartografie, dálkového průzkumu Země, statistiky, GIS a dalších geoinformačních disciplín. Nabízí:

- bakalářské studium oboru *Geoinformatika a kartografie*,
- navazující magisterské studium oboru *Geoinformatika*,
- navazující magisterské studium *Master's Degree Programme 'Copernicus Master in Digital Earth'* v rámci programu Erasmus Mundus Joint Master Degree (EMJMD), vyučovaný ve spolupráci s University of Salzburg a University of South Brittany,
- doktorské studium *Geoinformatika a kartografie*,
- doktorské studium *Geoinformatics and Cartography* v anglickém jazyce,
- rigorózní řízení v oboru *Geoinformatika a Geoinformatika a kartografie*.

Vědecko-výzkumná činnost katedry je zaměřena několika směry. Mezi hlavní směry lze zařadit obecné otázky geoinformatiky, budování a aplikace geografických informačních systémů, zpracování dat z dálkového průzkumu Země, tematickou a digitální kartografii, modelování prostorových jevů v krajině a další (Katedra geoinformatiky, 2021).

V akademickém roce 2020/2021 na katedře působilo 10 akademických pracovníků, jeden vědecký pracovník, sedm interních doktorandů a tři externí doktorandi. Bakalářské i magisterské studium každoročně úspěšně absolvují v průměru nižší desítky studentů. Jak uvádí Šabatová (2019), roční počty studujících jsou vázány i na demografický vývoj populace České republiky, což se posledních letech projevilo snižujícím se počtem studentů.

Katedra je členem, či má své zástupce, v mnoha odborných skupinách a organizacích (Česká asociace pro geoinformace, Česká kartografická společnost, Mezinárodní kartografická asociace, Nadace Open Source Geospatial Foundation, Společnost pro fotogrametrii a dálkový průzkum a další). Je partnerem řady českých i zahraničních vysokých škol. Svým studentům i učitelům nabízí atraktivní studijní pobyty z programů ERASMUS a CEEPUS. Za 20 let působení katedru navštívily desítky zahraničních hostů. Studenti i pracovníci katedry bývají za své akademické práce pravidelně oceňováni v prestižních soutěžích.

3.2.1 Popularizace a propagace katedry

Za účelem popularizace oboru se Katedra geoinformatiky každoročně účastní celosvětové oslavy „dne GIS“, jehož cílem je prezentovat a přiblížit geoinformační technologie veřejnosti. K této iniciativě se připojuje pořádáním akce *GIS Day*, určené především pro střední školy. Součástí programu akce obvykle bývají bloky krátkých přednášek, praktických cvičení, či ukázky technologií. *GIS Day* se mezi středními školami těší značné oblíbenosti, díky čemuž v současné době i přes vynikající organizační zabezpečení nestíhá uspokojovat veškerou poptávku (Voženílek, 2016).

Další popularizačně-propagační akcí pořádanou Katedrou geoinformatiky UP je *Víkend v Olomouci s geoinformatikou*. Jedná se blokový dvoudenní program určený zejména pro přihlášené uchazeče, ale i pro další studenty středních škol a veřejnost. Oproti *GIS Day* je program akce zaměřen více na jednotlivce a orientován na budoucí studenty. Cílem je představení olomoucké geoinformatiky jako studijního oboru a nastínění studia v Olomouci obecně. Součástí programu jsou ukázky technologií, či praktická cvičení.

Kromě vlastních propagačních akcí se katedra účastní i mnoha dalších. Mezi nejvýznamnější lze zařadit dny otevřených dveří fakulty i univerzity, *Noc vědců*, *Veletrh vědy*, dříve *Přírodovědný jarmark*, přednáškovou činnost v rámci Univerzity třetího věku či stylizovaným divadelním představením na Univerzitě dětského věku.

Mezi významnými propagačními činnostmi katedry nelze opomenout aktivity na sociálních sítích YouTube, Facebook a Instagram. Dalšími aktivitami jsou například série podcastů *GeoTalks* či živě vysílané cestovní přednášky o zahraničních výjezdech akademických pracovníků a studentů *KGI Talk*. Na poli medializace byli členové katedry několikrát hosty pořadů televizního vysílání (*Milénium*, *Dobré ráno*, *Reportéři ČT* a další) a rozhlasového vysílání (*Český rozhlas*, *Radiožurnál* a další). Různé aktivity katedry byly rovněž několikrát předmětem zpravodajství i na mnohých internetových portálech (*Seznam zprávy*, *Olomoucký deník*, *iDnes.cz*, *Olomoucká drbna* a další).

V rámci tištěných propagačních materiálů vznikla na katedře řada brožur a letáků, prezentujících studium geoinformatiky jako celek (brožura *Proč studovat geoinformatiku*, brožura *Studium geoinformatiky*), jednotlivé formy studia (brožura *PhD study*, leták *Dálkové navazující magisterské studium*), směry vědecko-výzkumné činnosti (leták *Eye-tracking*, leták *Remote Sensing*, leták *3D printing*), uplatnění absolventů v praxi (brožura *Jak se žije geoinformatikům*, brožura *UPsolventi geoinformatiky*) a další.

3.3 Data, informace a znalosti

S nástupem výpočetní techniky bezesporu začala nová éra – digitální. Společně s rozvojem výpočetní techniky rapidně akcelerovala i rychlost vzniku, sběru a evidence dat. V dnešním světě data vznikají závratnou rychlostí. Podle studie International Data Corporation (2017) bylo k roku 2016 v globální datové sféře vygenerováno a uloženo celkem 16,1 zettabytů dat (1 zettabyte ~ 10¹² gigabytů). Předpokládaný nárůst do roku 2025 je odhadován na 163 zettabytů. Jsou-li tato čísla převedena na roční procentní míru růstu, objem celosvětově uložených dat vzroste každý rok o 29,3 %. Očekávaný nárůst bude pravděpodobně ještě umocněn nově nastupujícími technologiemi jako internet věcí, umělá inteligence, autonomní a mobilní systémy a podobně.

Spolu s nárůstem uložených dat roste i potřeba se dále zabývat jejich ukládáním, strukturou, ale i analýzou. Data sama o sobě nenesou žádný význam, proto je potřeba v nich bádát, zkoumat je a nalézat v nich vzory a souvislosti. Teprve tak mohou mít

nashromážděná data skutečný přínos, být užitečná a přinášet cenné odpovědi, sloužící například pro podporu rozhodování nebo strategické plánování. V těchto souvislostech se stále více skloňují obory jako teorie informací, management informací, management znalostí, datová analytika a podobně. Souhrnně jsou tyto obory označovány termínem informační vědy.

3.3.1 Datová pyramida

V literatuře se v kontextu s pojmem data objevuje několik úzce souvisejících termínů. Do těchto na první pohled analogických, avšak významově rozdílných termínů jsou řazeny pojmy data, informace a znalosti. Někteří autoři navíc uvádějí pojmy porozumění a moudrost. Každý z pojmů představuje jinou úroveň abstrakce. Je nezbytné tyto termíny definovat a správně porozumět jejich vzájemným vztahům.

Data (angl. *data*) definoval Russel Ackoff, průkopník a přední odborník v oblasti teorií systémů, jako údaje reprezentující vlastnosti objektů a jevů (1989). Podle Bellingera a kol. (2004) jsou data definována jako surové údaje bez významu. Samy o sobě nenesou žádný smysl. Mohou existovat v jakékoliv podobě, užitečné i neužitečné. Podle Firestona (2000) jsou data hodnotou pozorovatelné, měřitelné nebo spočítatelné vlastnosti. Daty se obecně rozumí objektivní fakta o nějakém objektu nebo jevu, vyjádřená většinou číselnou nebo textovou hodnotou.

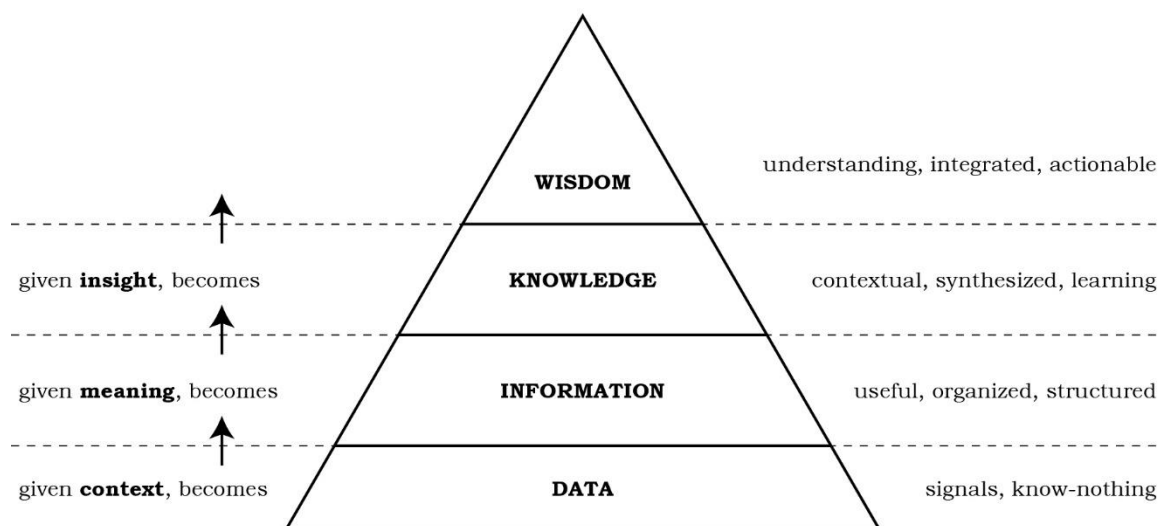
Informace (angl. *information*) jsou definovány jako data zpracovaná do ucelené podoby. Takto zpracovaná data již jsou sama o sobě nositelem významu. Význam je určen vztahy mezi zaznamenanými daty. Samotný význam pak může, ale i nemusí být užitečný. Informace běžně odpovídají na otázky typu kdo, co, kdy, kde a kolik (Bellinger a kol., 2004)

Znalosti (angl. *knowledge*) jsou naproti tomu označovány jako podmnožina informací, jejímž smyslem je nést význam užitečný. Jinými slovy se jedná o informace s přidanou hodnotou. Rozumí se jimi speciální typ informací, které byly podrobeny a prošly validačním testem, založeným buď na selském rozumu nebo ověření vědeckou hypotézou. V závislosti na typu testu jsou znalosti děleny na selský rozum a vědecké znalosti (Firestone, 2000). Jak uvádí Ackoff, znalosti odpovídají na otázky typu jak. Definice znalostí se různí. Někteří autoři pod pojem znalost zahrnují i pojmy porozumění a moudrost, jiní je definují jako samostatné a významově rozdílné pojmy.

Porozumění (angl. *understanding*) je interpolační, pravděpodobnostní, kognitivní, analytický proces, kterým je možno odvodit nové znalosti od předchozích znalostí (Bellinger a kol., 2004). Rozdíl mezi porozuměním a znalostí lze zjednodušeně vysvětlit jako rozdíl mezi učením a pamatováním. Porozuměním lze odpovídat na otázku typu proč.

Moudrost (angl. *wisdom*) značí nejvyšší úroveň abstrakce. Jedná se o extrapolaci, nepravděpodobnostní, nedeterministický proces. Na rozdíl od předchozích čtyř úrovní se ptá na otázky, na které není vždy jednoznačná a lehce dosažitelná odpověď. Je podstatou filozofického zkoumání. Jedná se o proces, díky kterému lze rozlišit nebo posoudit dobré a zlé. Jak uvádí Bellinger, moudrost je zákonitě pouze vlastnost lidská, protože vyžaduje přítomnost duše, kterou počítače nikdy mít nemohou.

Uvedené pojmy vyjadřují jednotlivé úrovně abstrakce a tvoří patra tzv. *datové pyramidy* (též *informační, znalostní*). V anglické literatuře je pyramida často označována jako *DIK hierarchie* (Data, Information, Knowledge), případně *DIKW hierarchie* (Data, Information, Knowledge, Wisdom), ilustrováno na obrázku 3.



Obr. 3 Datová pyramida (zdroj: Marcin Gajzler; grafická úprava: autor).

3.3.2 Záznam a atribut

Datové sady jsou tvořeny záznamy a jejich atributy. Záznam lze v této souvislosti chápat jako objekt uložený do datové sady. V případě datové sady studentů je za objekt považován student, který je do datové sady uložen v podobě záznamu. Jeden student pak odpovídá jednomu záznamu. Pakliže jsou záznamy ukládány do tabulky, jeden záznam je reprezentován jedním řádkem tabulky.

Uložený objekt je v datové sadě charakterizován atributy. Atributy jsou chápány jako popisné vlastnosti objektů. Pohledem statistiky je atribut nazýván náhodnou veličinou. Sada atributů popisující vlastnosti daného objektu je označována jako náhodný vektor. V případě ukládání záznamů do tabulky je atribut vyjádřen sloupcem. Záznamy a atributy dohromady tvoří tzv. *datovou matici*.

Atribut (vlastnost objektu) může být vyjádřen mnoha způsoby – textovým popisem, číslem, dichotomickou hodnotou, datumem nebo dokonce obrazovým či multimediálním záznamem. Způsob, jakým jsou data zaznamenána, je označován jako typ atributu (Han a kol., 2012).

3.3.3 Typy atributů

Atributy lze dělit z mnoha hledisek. Jak uvádí Šarmanová (2012), základní hlediska dělení dat jsou hledisko syntaktické a hledisko sémantické. Zatímco ze syntaktického se data dělí podle standardních datových typů (numerická, textová, logická, datumová atd.), ze sémantického hlediska se dále zabývá pouze daty strukturovanými a typy jsou děleny na kvalitativní a kvantitativní. Mezi kvalitativní jsou řazeny atributy nominální a ordinální. Mezi kvantitativní (též numerické) jsou řazeny atributy intervalové a poměrové.

Znalost atributového typu je stěžejním předpokladem k použití správné statistické metody. Použití nesprávné statistické metody vzhledem ke studovanému typu dat vede k chybné interpretaci dat a nepřesnému vyhodnocení.

Nominální

Nominální atributy (též kategoriální, kvalitativní) nabývají hodnot malého konečného počtu hodnot a znamenají příslušnost k jisté kategorii (Šarmanová, 2012). Hodnoty nelze uspořádat či seřadit do logické posloupnosti. V informačních vědách jsou nominální data nazývána také jako výčty.

Speciálním typem nominálního atributu je atribut binární. Binární atribut může nabývat pouze dvou kategorií. Kategorie jsou často reprezentovány hodnotami 0/1, ano/ne, pravda/nepravda. Pokud binární atribut nabývá hodnot pravda/nepravda, je označován jako *booleovský*. Binární atribut lze dále dělit na symetrický (obě hodnoty mají stejnou váhu) a asymetrický (jedna z hodnot má větší váhu, vyskytuje se vzácněji, druhá má spíše výchozí charakter).

Ordinální

Ordinální atributy jsou takové atributy, které lze seřadit do smysluplné logické souvislosti. Mezi uspořádanými hodnotami ale nelze určovat vzdálenosti. U ordinálních atributů je možné určit relativní pořadí hodnot, ale není možné vzdálenosti mezi jednotlivými hodnotami kvantifikovat.

Často se lze setkat s nominálními či ordinálními atributy vyjádřenými číselnou hodnotou. Číselná hodnota však slouží pouze jako zjednodušený zápis příslušné kategorie. Nelze ji v tomto případě používat ke kvantitativním měřením či operacím (např. určování průměru).

Intervalové

Intervalové atributy mohou nabývat hodnot, které lze řadit, a lze je mezi sebou kvantitativně porovnávat. Hodnoty mají konstantní a spočítatelné měřítko. Klíčovou vlastností intervalových hodnot je možnost určit vzdálenost mezi dvěma hodnotami. Nedostatkem intervalových dat je chybějící počátek číselné osy. Projevem absence počátku číselné osy je, že nejsme schopni určit nulovou hodnotu, a tedy nejsme schopni aplikovat značné množství statistických metod. Příkladem intervalových dat je teplota.

Poměrové

Poměrové atributy mohou podobně jako intervalové atributy nabývat reálných hodnot, mezi kterými je možné kvantifikovat vzdálenosti. Narozdíl od intervalových je také možné určit počátek číselné osy (absolutní nulu). To znamená, že v tomto případě lze o naměřené hodnotě mluvit jako o násobku (poměru) jiné naměřené hodnoty (Han a kol., 2012). Příkladem poměrových atributů může být vzdálenost, váha nebo výška.

3.3.4 Způsoby uložení dat

Nezbytnou součástí procesu analýzy dat je management datových zdrojů. Jeho kvalita zákonitě předurčuje plynulost celého procesu. Do oblasti managementu dat spadá i způsob, jakým jsou data uložena a spravována. Způsob uložení je závislý zejména na povaze dat a zvolené analytické metodě, kterou je zamýšleno pro analýzu použít.

Souborový systém

Nejstarším a nejjednodušším způsobem je uložení dat v souborovém formátu. Jedná se o prosté uložení souboru na disk nebo úložiště. Tento způsob poskytuje flexibilitu v ukládání různých datových formátů, zásadní nevýhodou je ale obtížné dotazování a vyhledávání v datech. Souborový systém je vhodné používat spíše pro relativně malé datové sady. Z pohledu dnešních databázově-orientovaných úložišť je tento typ označován jako prostý databázový soubor (angl. *flat file*).

Často je možné se setkat se zápisem dat v tzv. *delimitovaných formátech*. Jedná se o způsob zápisu strukturovaných dat ve formě matice či tabulky, v níž jsou hodnoty odděleny specifickým, předem stanoveným symbolem. První řádek matice definuje její strukturu, další řádky pak uchovávají hodnoty záznamů. Typickým příkladem tohoto typu souboru je formát CSV. Dalším způsobem zápisu strukturovaných dat jsou značkovací formáty, například XML nebo JSON.

Speciálním typem souboru jsou soubory tabulkových procesorů. V tomto případě se jedná o aplikačně specifické formáty jako XLS (Excel Binary File Format) nebo ODS (Open Document Spreadsheet). Takové formáty kromě samotných dat nesou například funkce a další metadata.

Databáze

Pokročilejším způsobem uložení dat je uchování v databázi. Uchování dat v databázi umožňuje jednoduše k nim přistupovat, aktualizovat a odstraňovat je. Databáze může sloužit k ukládání strukturovaných dat (relační databáze), ale i nestrukturovaných dat (tzv. *NoSQL* databáze).

V dnešní době je nejpoužívanějším typem databáze relační. Relační databáze strukturovaných dat jsou založeny na uchovávání informací v předem definované struktuře skládající se z tabulek a vazeb (relací) mezi nimi. Vazby jsou realizovány pomocí tzv. klíčů. Vázáním tabulek mezi sebe lze docílit snížení redundance, což má za následek například vylepšení výkonu databáze (North, 2012). Nespornou výhodou je možnost se na data dotazovat skrze systém řízení báze dat pomocí dotazovacích jazyků (například SQL).

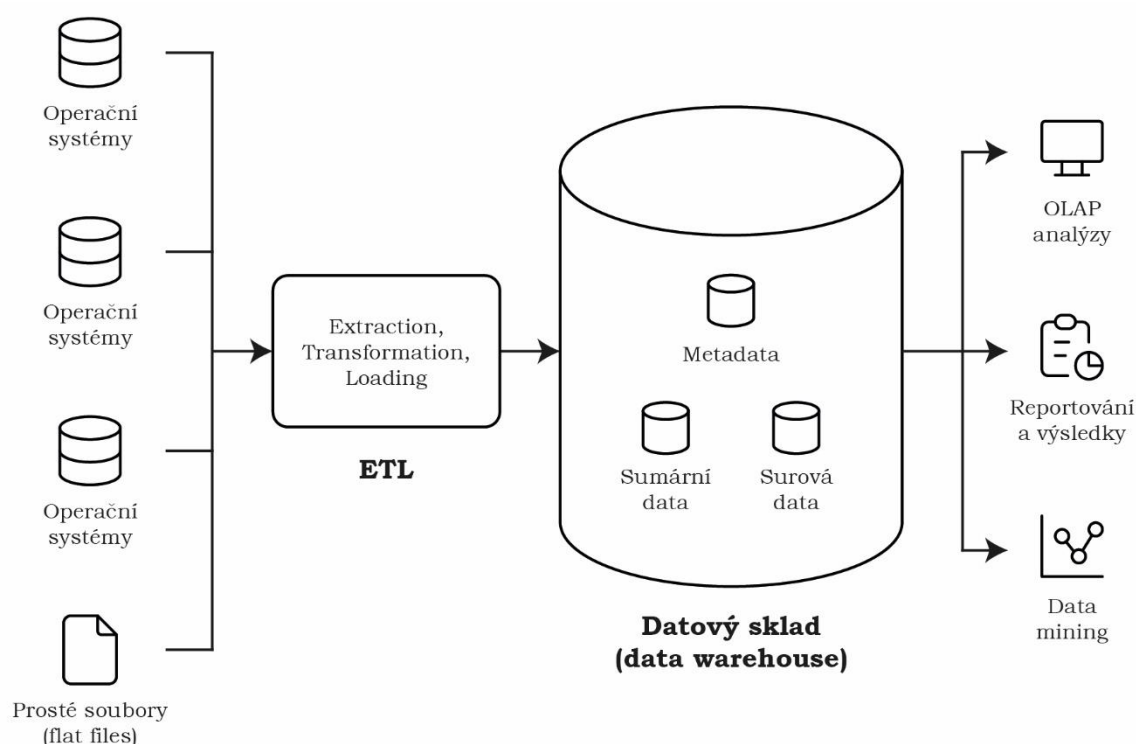
Většina relačních databází je dimenzována na velký počet zápisů a čtení. Z tohoto důvodu jsou označovány jako transakční (též operační) systémy (angl. *Online Transaction Processing, OLTP*). Příkladem transakční databáze jsou bankovní systémy registrující platby, obchodní systémy registrující nákupy apod. Tyto systémy jsou velmi efektivní v případě obrovského množství rychle probíhajících transakcí. Pokud je potřeba nad takovými daty provádět analýzy, efektivita databáze prudce klesá. Pro získání agregovaných dat by bylo nutné vytvořit dotaz kombinující více tabulek. V případě tabulek o milionech záznamů by se jednalo o výpočetně náročný a časově zdoluhavý proces.

Řešením je vytvoření tabulek obsahujících denormalizovaná či agregovaná data, generovaná ze surových transakčních dat. Nad takto vytvořenými daty pak lze analýzy provádět efektivněji. Databázové systémy navržené primárně pro analýzy a reporting se nazývají analytické (angl. *Online Analytical Processing, OLAP*). Proces převedení z OLTP systému do OLAP systému je v literatuře označován zkratkou *ETL* (Extraction, Transformation, Load).

Datový sklad

Datový sklad (angl. *data warehouse*) definoval W. H. Inmon jako podnikově strukturovaný depozitář subjektivě orientovaných, integrovaných, časově proměnlivých, historických dat použitých na získávání informací a podporu rozhodování. V datovém skladu jsou uložena atomická i sumární data (Lacko, 2003). Data uložená v datovém skladu mohou sloužit mnoha různým účelům, včetně pasivního čekání na jejich budoucí potřebu (Inmon, 2005). Datový sklad je často spravován zcela odděleně od podnikových operačních databází. Aby mohl datový sklad sloužit pro podporu rozhodování, musí adoptovat principy OLAP systémů, a tedy splňovat funkční a výkonnostní požadavky, kterými OLTP systémy nedisponují (Chaudhuri a kol., 1997).

Datový sklad je v podstatě systém relačních databází obsahujících pouze data pro podporu rozhodování. Taková data jsou z transakčních databází metodami ETL extrahována, pečlivě selektována, transformována, agregována a následně ukládána do datového skladu. Ze skladu je pak možné data vybírat, analyzovat, vytvářet reporty a vizualizovat (obr. 4)



Obr. 4 Data warehousing (zdroj: IBM; grafická úprava: autor).

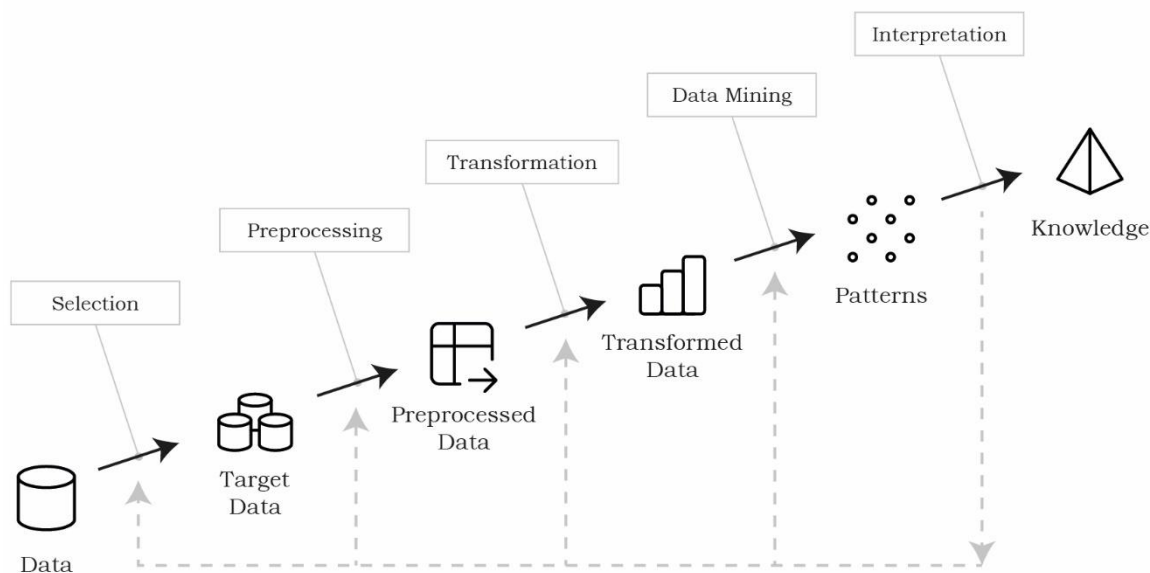
3.4 Datová analýza

Analýzou se obecně rozumí rozbor, zejména ve smyslu všeobecné metody zkoumání jednotlivých složek a vlastností nějakého předmětu, jevu nebo činnosti. Slovník Merriam-Webster (2022) definuje analýzu jako detailní zkoumání celku se záměrem pochopit jeho podstatu nebo určit jeho základní vlastnosti. Cílem analýzy je dekomponovat celek na elementární části, identifikovat podstatné vlastnosti částí a poznat jejich zákonitosti. Akt analýzy lze v obecnějším smyslu chápat i jako podrobné studium nebo zkoumání jevu s cílem zjistit o daném jevu více (Cambridge Dictionary, 2022). Analytikou se pak nazývá metoda logické analýzy a obor založený na analytické metodě, používající metody rozboru (Ústav pro jazyk český, 2021).

Obor datové analytiky definuje Runkler (2020) jako aplikaci analýz na rozsáhlé datové sady s využitím výpočetní techniky. Cílem datové analýzy je zejména získat podklady pro podporu rozhodování. Jedná se o velmi rozmanitý a interdisciplinární vědní obor, který převzal principy z mnoha dalších vědních disciplín jako např. statistika, strojové učení, teorie systémů nebo umělá inteligence.

Možnosti oblastí využití datové analýzy jsou téměř nekonečné. Datová analytika již není doménou pouze akademickou, ale čím dál více nachází uplatnění v kruzích komerční sféry, strojírenství, medicíny atd. Jako příklad aplikace datové analytiky lze uvést oblasti jako optimalizace výroby, segmentace zákazníků nebo klinické studie.

Jak uvádí Dobešová (2004), datová základna sama o sobě je jen pouhé skladiště dat. Nové informace pro řízení a rozhodování jsou získány až aplikací zkušeností a znalostí na tato data. Podle existence či neexistence fyzických vazeb ve zkoumaných datech lze analytický přístup rozdělit na dvě kategorie. První kategorií je tzv. *OLAP* (viz kapitola 3.3.4 *Způsoby uložení dat*). Druhou kategorií je *data mining* (obr. 5). Ten lze obecně charakterizovat jako proces extrakce relevantních, předem neznámých nebo nedefinovaných proměnných informací z velmi rozsáhlých databází (Dobešová, 2004). Cílem práce je použít právě metody dolování z dat a získat tím nové informace o uchazečích, studentech a absolventech Katedry geoinformatiky UP.



Obr. 5 Proces dobývání znalostí z databází (zdroj: Fayyad a kol.; grafická úprava: autor).

4 DOSTUPNÉ DATOVÉ ZDROJE

Byla provedena rešerše existujících dat spojených s Katedrou geoinformatiky UP, či dat potenciálně přínosných v analýze. Výsledkem rešerše je celá řada datových sad. Sady související s Katedrou geoinformatiky UP mají často původ v akademické činnosti, kdy jsou sbírána v rámci bakalářských a magisterských prací formou dotazníků. Dalším typickým zdrojem jsou agendy a rejstříky vedené za evidenčními účely. V některých případech jsou souhrnem a agregací vycházející z jiných dohledaných datových sad.

4.1 Dotazníkové šetření Víchové

Sběrem dat o uchazečích, studentech a absolventech Katedry geoinformatiky UP se ve své práci zabývala Víchová (2007). Pro šest až sedm předem stanovených bodů na časové ose každého studenta sestavila dotazník (pro uchazeče, pro studenty po ukončení prvního semestru, pro studenty po ukončení třetího semestru atd.). Obsah každého dotazníku byl přizpůsoben bodu na časové ose (dotazník pro uchazeče orientován na povědomí o studiu a motivaci, dotazník pro absolventy orientován na zaměstnání atd.). Každý respondent v závislosti na aktuální pozici v rámci své časové osy vyplnil příslušné dotazníky, z čehož vyplývá, že podstatná část dotazníků byla vyplněna zpětně. Jak uvádí Víchová, zpětný sběr dotazníků má za následek jejich sníženou vypovídací hodnotu.

Tímto způsobem byla získána data o 333 uchazečích a 326 studentech od akademického roku 1997/1998 do 2006/2007. Přestože výstupní databáze Víchové obsahuje tabulku s vytvořeným datovým modelem pro ukládání absolventů, jejich záznamy se nepodařilo dohledat. Ostatní data získaná z dotazníkového šetření uložila do relační databáze. Šabatová (2019) uvádí, že na práci bylo navázáno projektem, v rámci kterého byla data o uchazečích, studentech a absolventech pomocí dotazníků sestavených Víchovou sbírána až do roku 2012.

4.2 Dotazníkové šetření Šabatové

Na práci Víchové (2007) navázala Šabatová (2019), přičemž data sbírala prostřednictvím nově vytvořeného dotazníkového šetření. Šetření bylo provedeno online, rozděleno na tři samostatné části podle cílové skupiny (uchazeči, studenti, absolventi) a jednotlivé dotazníky jim byly rovněž obsahově přizpůsobeny. Celkem získala data o 36 uchazečích, 95 studentech a 174 absolventech. Data uložila do tří samostatných datových matic podle typu respondenta.

Podobně jako Víchová zařadila Šabatová do dotazníků množství otevřených otázek, týkajících se například zážitků ze zahraničních výjezdů, oblíbeného místa trávení volného času během studia, či zaměstnání po absolutoriu. Odpovědi na otevřené otázky dále manuálně třídila do stanovených kategorií. Šabatová uvádí, že ne vždy bylo možné odpověď jednoznačně vyhodnotit. Vhodnost otázek Šabatová konzultovala s pracovníky Katedry geoinformatiky UP.

Přestože dotazníky Víchové a Šabatové se ve značném počtu otázek liší, v některých lze pozorovat podobnost. Této skutečnosti lze případně využít při následném kombinování datových sad. Prostřednictvím párování totožných otázek lze vytvořit rozšířenou výchozí datovou sadu a docílit tak vyšší vypovídací hodnoty zjištěných závěrů. Současně je však nezbytné zamezit vzniku duplicitních záznamů zapříčiněných odpovědi respondenta v obou dotazníkových šetřeních.

4.3 Studijní agenda Univerzity Palackého (STAG)

V rámci administrace studijní agendy Univerzity Palackého v Olomouci jsou data o studentech evidována a uchovávána v informačním systému IS/STAG. Systém umožňuje evidovat studenty prezenční i kombinované formy studia, studenty celoživotního vzdělávání i účastníky univerzity třetího věku (Západočeská univerzita v Plzni, 2021). V systému jsou dále vedena i data o uchazečích o studium. Na žádost byla data poskytnuta Studijním oddělením Přírodovědecké fakulty Univerzity Palackého v Olomouci, zastoupeným Mgr. Jiřím Mazalem.

O studentovi je vedeno množství popisných charakteristik, které lze rozdělit do několika tematických kategorií (demografické údaje, údaje o absolvované střední škole, údaje o studovaném oboru, údaje o průběhu studia, prospěch studenta v minulých letech, způsob ukončení studia, identifikátory a další). O uchazeči jsou dále vedena data o přijímací zkoušce a zápisu ke studiu. Data jsou v podobě tabulek dostupná vždy za akademický rok.

Z výše uvedeného je patrná důvěrnost a citlivost některých evidovaných údajů. Nezbytným předpokladem pro obdržení dat je uzavření dohody o mlčenlivosti mezi organizací poskytující data a jejich zpracovatelem. Data či jakékoliv další informace z dat vyplývající je dále možné poskytovat či prezentovat pouze v anonymizované podobě tak, aby nedošlo k porušení podmínek stanovených dohodou o mlčenlivosti.

Jak uvádí Šabatová (2019), protože Studijní agenda byla původně vedena papírovou formou, nelze získat ucelená data za celé období působení Katedry geoinformatiky UP. V elektronické podobě jsou data o studentech dostupná nejdříve od akademického roku 2000/2001 (včetně) a data o uchazečích od 2003/2004 (včetně). Data jsou dostupná v jednoletém intervalu akademického roku až do 2021/2022 (včetně).

Je důležité zmínit, že data za akademický rok lze považovat za finální až po termínu uzavření akademického roku. Tento fakt má v datech aktuálně probíhajícího roku za následek nízkou vypovídací hodnotu některých evidovaných atributů (např. počty kreditů, studijní průměr). Jinými slovy, data z neuzavřeného akademického roku nelze považovat za kompletní, na což je třeba dbát při dalším zpracování. Řešením je například práce pouze s uzavřenými roky nebo zahrnutí dat z neuzavřených let v omezeném rozsahu (tj. výběr pouze vypovídajících atributů).

4.4 Interní data Katedry geoinformatiky UP

Na Katedře geoinformatiky UP je dále interně vedeno množství datových sad určených pro akademické či evidenční účely. Data jsou většinou spravována jednotlivými členy katedry a případně na žádost a po osobní domluvě poskytována.

Jednou z vedených datových sad jsou data o zdroji povědomí uchazečů o Katedře geoinformatiky UP. Datový model obsahuje identifikační údaje uchazeče název střední školy, typ střední školy, a především informace o zdroji povědomí. Struktura dotazníku obsahuje 12 možných odpovědí na otázku „Jak jste se dozvěděli informace o studiu na Katedře geoinformatiky Univerzity Palackého v Olomouci“ (Dobešová, 2018). Mezi odpověďmi lze uvést „od učitele zeměpisu“, „od učitele informatiky“, „od kamaráda“ a další. Lze uvést i více možností současně. Obsahem je i informace, zda uchazeč nakonec ke studiu nastoupil. Dotazníky jsou od uchazečů získávány na přijímacích zkouškách, konzultačních dnech, akci GIS Day apod.

Datová sada je kontinuálně vedena od roku 2016 (uchazeči o studium v akademickém roce 2016/2017) do 2019 (uchazeči o studium v akademickém roce 2019/2020). V roce 2020 nebylo možné z důvodu nouzové situace způsobené pandemií COVID-19 data relevantně posbírat. Analýzou těchto dat se zabývali Dobešová a Piňos (2019). Data pro účely diplomové práce poskytla doc. Ing. Zdena Dobešová, Ph.D.

Interně je na Katedře geoinformatiky UP vedena datová sada účastníků mobilit. Sada byla historicky vedena manuálně. Od roku 2016 je sada vedena automatizovaně v aplikaci Erasmus Plus. Po spojení obou částí vznikla sada čítající celkem 251 záznamů s atributy jména a příjmení účastníka, destinací (stát, město, univerzita), role, v jaké účastník mobilitu absolvoval a programem mobility. Data pro účely diplomové práce poskytl doc. RNDr. Jaroslav Burian, Ph.D.

Od roku 2017 jsou shromažďovány reporty o praxích bakalářských a magisterských studentů. Reporty obsahují informace o firmách, ve kterých jsou praxe absolvovány, identifikační údaje studenta, časové označení, hodnocení praxí studenty a hodnocení studentů svými tutory. Data byla vzhledem ke své citlivé povaze poskytnuta v anonymizované podobě. Obsahem sady je celkem 150 záznamů absolventů praxí rozdělených do tří samostatných souborů. Data pro účely diplomové práce poskytl RNDr. Jan Brus, Ph.D.

Dále je pro evidenční účely vedena datová sada přijatých uchazečů. Data ale mají pouze anonymizovanou a agregovanou podobu za jednotlivé akademické roky. Obsahují informace o způsobu přijetí uchazeče ke studiu (na základě prospěchu na střední škole, na základě přijímací zkoušky, na základě odvolání atd.). Na základě totožnosti kódů rozhodnutí lze usuzovat, že tato datová sada je pravděpodobně přímo odvozená z dat Studijní agendy. Časový rozsah dat je od akademického roku 2003/2004 do roku 2007/2008.

Za relevantní data lze považovat i záznamy o popularizačních přednáškách studentů a zaměstnanců na středních školách (interně nazývané *aGITace*). Datový model obsahuje jméno a příjmení přednášejícího, název, místo a typ střední školy, časovou značku (rok), typ akce (domluvená prezentace, přednáška v rámci jiné akce). Data jsou dostupná od roku 2015.

4.5 Externí datové zdroje

Vhodnými daty k zahrnutí do analýz jsou data z *Veřejné databáze* poskytované Českým statistickým úřadem. Každý údaj je v databázi identifikován z hlediska věcného, časového i územního, a také s odkazem na zdroj dat (Český statistický úřad, 2022). Díky tomu lze v databázi velmi efektivně vyhledávat relevantní data a na základě vlastního výběru data exportovat do požadovaného formátu. Obsahem databáze je množství statistických ukazatelů indikujících stav a vývoj České republiky. Obzvláště přínosnými pro následné analýzy by pak mohly být zejména demografické charakteristiky (skupina *Obyvatelstvo*), dojíždka (skupina *Sčítání lidu, domů a bytů*), nebo další školská data (skupina *Vzdělávání*).

Uceleným a rozsáhlým souborem dat o vysokých školách je databáze *Sdružené informace matrik studentů* (zkráceně SIMS). Vysokým školám a poskytovatelům zahraničního vysokoškolského vzdělávání působícím na území České republiky je podle zákona č. 111/1998 Sb. uložena povinnost vést matriku studentů. Matrika dále slouží k účelům rozpočtovým a statistickým (Ministerstvo školství, mládeže a tělovýchovy,

2021). Každá vysoká škola a zahraniční poskytovatelé jsou dále povinni data ukládat do centrální databáze a trvale je aktualizovat (zpravidla čtyřikrát ročně). Tato data jsou v agregované souhrnné podobě poskytována veřejnosti na webu MŠMT. Údaje jsou dostupné od roku 2000.

V *Rejstříku škol a školských zařízení* jsou v podobě tabulky poskytovány údaje o školských zařízeních poskytujících základní a střední vzdělávání. Předmětem datové sady jsou zejména identifikační údaje školy (IZO, IČO, název atd.), adresa školy, kontaktní údaje školy apod. Data jsou poskytována Ministerstvem školství, mládeže a tělovýchovy a dostupná v aktuální podobě.

Tab. 1 Dostupné datové zdroje související s KGI UP *

Id	Název	Zdroj/správce	Časové určení
1	Studijní agenda uchazečů	Studijní odd. PrF UP	2003/04–2020/21
2	Studijní agenda studentů	Studijní odd. PrF UP	2000/01–2020/21
3	Katederní sada uchazečů	doc. Ing. Zdena Dobešová, Ph.D.	2016/17–2019/20
4	Katederní sada mobilit	doc. RNDr. Jaroslav Burian, Ph.D.	2004/05–2021/22
5	Katederní sada praxí	RNDr. Jan Brus, Ph.D.	2017/18–2021/22
6	Katederní sada aGITací	doc. Ing. Zdena Dobešová, Ph.D.	2015/16–2018/19
7	Dotazník pro uchazeče	Bc. Marie Šabatová	jaro 2019 (sběr dat)
10	Dotazník pro studenty	Bc. Marie Šabatová	jaro 2019 (sběr dat)
11	Dotazník pro absolventy	Bc. Marie Šabatová	jaro 2019 (sběr dat)
12	Dotazníkové šetření Vichové	Mgr. Zdeňka Vichová	2006 (sběr dat)

* *Data ze Studijní agendy a data z katederních sběrů dat jsou k dispozici i za další časová období, tabulka uvádí časové určení, které vstupovalo do analýz v rámci této diplomové práce.*

5 PŘÍPRAVA DAT

Data nashromážděná v první fázi práce byla v nezpracovaném surovém formátu a často v neucelené podobě. Protože kvalita zpracování analýz je přímo závislá na způsobu uložení a managementu dat, bylo nezbytné je uvést do takové podoby, která umožní hladký průběh celého analytické procesu. Za tímto účelem byla provedena řada operací vedoucích k vytvoření souboru základových datových sad vstupujících do analýz.

5.1 Výchozí datové sady

Z dostupných datových zdrojů popsaných v rešerši byly pro další analýzy vybrány právě sady s nejvyšším potenciálním informačním ziskem. Výběr proběhl na základě autorem stanovených kritérií, mezi které byla zařazena atributová složka dat, počet záznamů v datové sadě, úplnost sady z hlediska chybějících hodnot atributů nebo časový rozsah. Roli hrála taktéž atraktivita tematické složky a její potenciální přínos.

Tímto způsobem bylo pro analyzování zvoleno celkem sedm tematických datových sad ze tří různých zdrojů a vybrána jedna kandidátní sada pro případné doplnění tematických dat. Kompletní přehled vybraných dat je uveden v tabulce 2. Data byla uložena prostým souborovým způsobem (*flat file*) v delimitovaných formátech CSV a v aplikačních formátech XLS. Prvním krokem při zpracování bylo spojení souborů každé sady do jedné základové sady.

Tab. 2 Přehledová tabulka použitých dat *

Id	Název	Zdroj/správce	Časové určení
1	Studijní agenda uchazečů	Studijní odd. PřF UP	2003/04–2020/21
2	Studijní agenda studentů	Studijní odd. PřF UP	2000/01–2020/21
3	Katederní sada uchazečů	doc. Ing. Zdena Dobešová, Ph.D.	2016/17–2019/20
4	Katederní sada mobilit	doc. RNDr. Jaroslav Burian, Ph.D.	2004/05–2021/22
5	Dotazník pro uchazeče	Bc. Marie Šabatová	jaro 2019 (sběr dat)
6	Dotazník pro studenty	Bc. Marie Šabatová	jaro 2019 (sběr dat)
7	Dotazník pro absolventy	Bc. Marie Šabatová	jaro 2019 (sběr dat)

* Data ze Studijní agendy a data z katederních sběrů dat jsou k dispozici i za další časová období, tabulka uvádí časové určení, které vstupovalo do analýz v rámci této diplomové práce.

5.2 Spojení datových modelů

Data ze Studijní agendy byla za období 2000/01–2017/18 získána z práce *Analýza a vizualizace prostorové diferenciac studentů a absolventů Katedry geoinformatiky UP* (Šabatová, 2019). Tento rozsah dat byl aktualizován o záznamy za akademické roky 2018/19–2020/21. Data za toto období byla poskytnuta Studijním oddělením PřF UP. Datové modely obou částí ale nebyly shodné, k čemuž došlo v důsledku rozdílného způsobu exportu atributů. Datové sady se překrývaly zhruba v 70 % atributů. Atributy v rámci tohoto překryvu byly zachovány, neodpovídající atributy byly zcela odstraněny a z dalších analýz vynechány.

Obdobným způsobem byly propojeny dotazníkové sady Šabatové. Protože licenční plán online nástroje použitého pro sběr dotazníků neumožňoval všechny odpovědi respondentů posbírat na jeden uživatelský účet, účty byly použity dva. Z obou účtů byla data exportována a vytvořeny základové sady uchazečů, studentů a absolventů.

Slučování dat bylo nezbytné i v případě interních dat o mobilitách. V tomto případě se jednalo o sloučení manuálně vedené databáze a dat exportovaných z aplikace Erasmus Plus. Z důvodu nekonzistence datových modelů byl vytvořen takový model, pro který bylo možné získat maximum atributových záznamů. Vytvořený datový model byl následně manuálně naplněn dostupnými daty.

5.3 Geokódování

Prostorové informace se v datech nacházely vyjádřené nepřímo, ve formě adresních bodů. Typicky se jednalo o adresy trvalých a přechodných bydlišť, adresy středních škol nebo adresy pracovišť. Tyto geokódy byly prostřednictvím služby *Awesome Table* v procesoru *Google Sheets* převedeny na přímé určení ve formě zeměpisných souřadnic. Celkem bylo zpracováno přes 12 100 záznamů. Celý proces z důvodu denního limitu použitého nástroje probíhal na etapy v rozmezí několika dní.

5.4 Hrubá filtrace a čištění

Obsahem dat bylo kromě relevantních atributů i množství pro analýzy nepoužitelných a nadbytečných. Tyto zbytné atributy byly pro hladký průběh a přehlednost z datových sad odstraněny. Typicky se jednalo o časové značky odpovědí či data exportu v případě dotazníků a redundantní atributy v případě dat Studijní agendy. Dalším krokem bylo odstranění testovacích záznamů z dotazníku. Testovací záznamy se v sadách vyskytovaly pouze v malém množství (jednotky) a proto mohly být odstraněny manuálně.

5.5 Transformace

Transformací dat v tomto kontextu se rozumí uvedení hodnot záznamů do takové podoby, která umožňuje provedení analýzy bez nutnosti dalšího destruktivního zásahu do dat. Příkladem operace transformace je parsování atributu ročních prospěchů studentů. Hodnoty tohoto atributu byly v datech reprezentovány jednou buňkou pro každý záznam, uvnitř které se nacházely roční průměry studenta za celé studium. Průměry byly uvedeny v číselném formátu delimitovaném středníky. Parsování proběhlo v prostředí *RapidMiner Studio*. Výsledkem byl vznik samostatného sloupce pro každý roční průměr. Dalšími operacemi bylo nastavení odpovídajících typů atributů (binární hodnota, číslo, datum atd.), harmonizace názvů atributů, dichotomizace některých výčtových atributů, převod hodnot do požadovaných jednotek atd.

Zmíněnými kroky bylo připraveno sedm základových datových sad. Sady byly uvedeny do stavu splňujícího všechny předpoklady pro hladký průběh analýzy. Zpětně lze tuto přípravu považovat za velmi opodstatněnou a přínosnou. Během analýz nebyly zjištěny žádné nedostatky a nebylo nutné v primárních surových datech zpětně provádět žádné další úpravy. Veškeré analýzy vycházely z takto připravených dat.

6 ZÁKLADNÍ STATISTICKÁ ANALÝZA

Jako první krok proběhlo pečlivé seznámení se s analyzovaným materiálem. Data byla nejprve zkoumána vizuálně, jednotlivé atributy byly zasazeny do kontextu celku a byly stanoveny kandidátní témata pro analýzu. Vybudování celkového přehledu o datech výrazně dopomohlo hladkému průběhu analytického procesu.

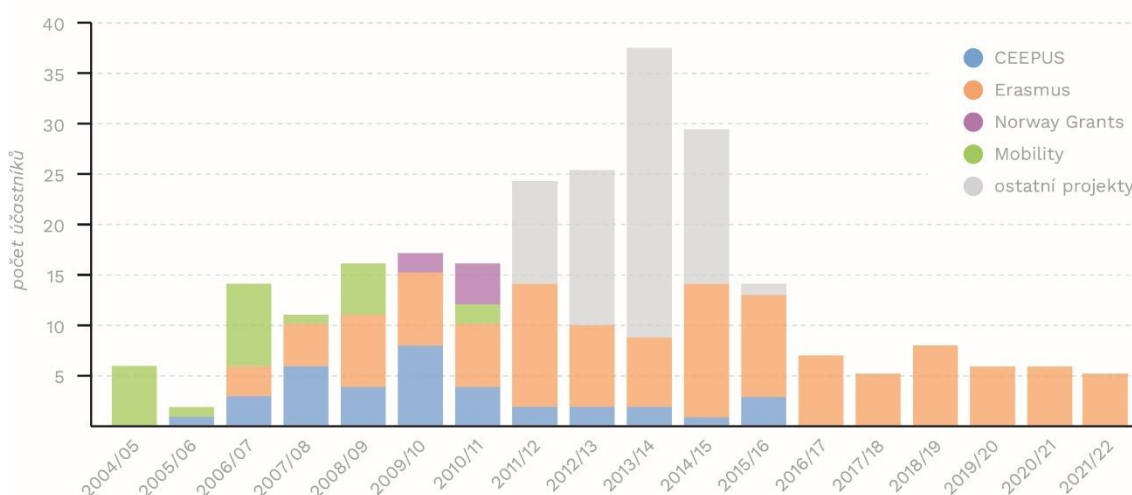
6.1 Statistické infografiky

V prostředí *RapidMiner Studio* nad sadami proběhla základní statistická analýza. Pomocí kontingenčních tabulek zpracovaných nad atributy byly vytvořeny jednoduché sumarizační grafy. Rovněž byly vypracovány základní mapové výstupy prezentující prostorovou diferenciaci záznamů, většinou jednoduchou bodovou metodou nebo metodou kartodiagramu. Mapové výstupy byly připraveny v *ArcGIS for Desktop*.

Oba typy výstupů byly v surovém vektorovém formátu ze zmíněných prostředí exportovány a do výsledné grafické podoby uvedeny v softwaru *Adobe Illustrator*. Typickými úpravami bylo přebarvení grafů do definované barevné palety, sjednocení stylů grafů a úprava kompozice či její doplnění (titul, podtitul, legenda) atp. (obr. 6)

D.6.3 VÝVOJ POČTU ÚČASTNÍKŮ MOBILIT PODLE PROGRAMU

v akademických letech 2004/05–2021/22



Obr. 6 Ukázka vybraného grafického výstupu.

Grafické výstupy statistické analýzy byly zpracovány do podoby přehledových infografik reprezentujících datové sady. Pro každou sadu byla vytvořena samostatná infografika obsahující základní informace o sadě (časový rozsah, počet záznamů, počet unikátních záznamů), stručný popis, grafy nejdůležitějších atributů a mapový výstup. Sestavené grafické výstupy lze dohledat v příloze 9. *Webová prezentace* v sekci *Pokročilá analýza*.

Tvorba výsledných vizualizací probíhala v softwaru *Adobe InDesign*. Infografiky jsou vyvedeny v jednotné, logicky uspořádané kompozici. Každá infografika je charakterizována barevným schématem příslušným k danému zdroji dat. Primárně jsou výstupy koncipovány pro digitální zobrazení na obrazovce (prezentaci na webu), avšak s vědomím potenciálně možného tisku zpracování proběhlo v barevném modelu CMYK, v rozměrech formátu A3 (420×297 mm) a rozlišení 300 dpi.

6.2 Výsledky statistické analýzy

Přestože statistické zpracování bylo provedeno prvotně za účelem vybudování základního přehledu o obsahu výchozího materiálu, již to samotné přineslo mnohá zajímavá fakta. Nejzajímavější zjištění a potvrzení předpokládaných skutečností jsou uvedena v následujících odstavcích. Odkazované grafické výstupy lze dohledat v příloze 9. *Webová prezentace* v sekci *Pokročilá analýza*.

Studijní agenda uchazečů (STAG_UCHAZEC)

Z celkového počtu uchazečů o studium na KGI bylo přijato 69,5 % (výstup D.1.3). Z toho se ke studiu zapsalo 65,2 % přijatých (výstup D.1.4). Zhruba každý třetí přijatý uchazeč se tedy ke studiu nezapiše. Podle očekávání lze z prostorové diferenciaci uchazečů v rámci České republiky nejvyšší koncentraci pozorovat v moravských regionech, nejvíce v Olomouckém a Zlínském kraji, odkud pochází nejvíce uchazečů (výstup D.1.5). Z grafu vývoje počtu hlásících se uchazečů je zřetelný sestupný trend v posledních letech, který je pravděpodobně dán celorepublikovým demografickým vývojem (výstup D.1.6). Tento závěr je v souladu s tvrzením Šabatové (2019).

Studijní agenda studentů (STAG_STUDENT)

Na histogramu celkového studijního průměru lze nejvyšší zastoupení pozorovat v intervalu 1,75–2,00 (výstup D.2.2). Z hlediska celkového rozložení četností v intervalech je graf učebnicovým příkladem Gaussova rozdělení. Z grafu studentů podle ukončení studia vyplývá skutečnost, že pouze 4,3 % studentů studium absolvoje s vyznamenáním (výstup D.2.4). Anomálií je fakt, že v druhém ročníku navazujícího magisterského studia průměrně studuje o 0,6 studentů více studentů než v ročníku prvním (výstup D.2.6). Anomálie je způsobeno tím, že v navazujícím *Master's Degree Programme 'Copernicus Master in Digital Earth'* v rámci programu *Erasmus Mundus Joint Master Degree* nastupují studenti na Univerzitu Palackého v Olomouci až do druhého ročníku, když první rok studují na University of Salzburg.

Dotazník pro uchazeče (DOT_UCHAZEC)

Dotazníková sada pro uchazeče je z hlediska počtu záznamů skromná. Odpovědělo celkem 33 respondentů hlásících se do akademického roku 2019/2020. Více než 65 % z nich neuvedlo identifikační údaje (výstup D.3.4), což znemožní plnohodnotné propojení sady s dalšími záznamy ze sad jiných. Za zmínku stojí statistika představy o budoucím povolání, podle níž téměř 72 % uchazečů nemá představu o kariéře po absolvování studia (výstup D.3.3). V případě existence tohoto atributu v jiných sadách bude zajímavé sledovat jeho vývoj. Nejčastěji uváděným zdrojem o povědomí o studiu je podle očekávání internet, na druhém místě pak veletrh vysokých škol Gaudeamus.

Dotazník pro studenty (DOT_STUDENT)

Podle odpovědí respondentů vycestuje na zahraniční mobilitu zhruba každý pátý (výstup D.4.2). Většina studentů tráví na katedře v průměru 15–20 hodin týdně (výstup D.4.3). Nejvíce studentů (77,3 %) chce po absolvování pracovat v komerční sféře (výstup D.4.6). Z hlediska pracovního zaměření je situace pestřejší a skupiny jsou rovnoměrněji rozložené do jednotlivých odvětví (výstup D.4.7). V rámci Olomouce nejvíce studentů bydlí v okolí centra města, nejvíce obývanou částí je Olomouc–město (výstup D.4.5).

Dotazník pro absolventy (DOT_ABSOLVENT)

Většina zúčastněných respondentů je absolventem magisterského studia (výstup D.5.1). Při porovnání grafů sféry uplatnění po studiu a grafu představy o budoucím povolání během studia lze pozorovat rozdíly v podílech zastoupení v komerční a státní sféře (výstup D.5.3). Pakliže se budou považovat preference studentů v tomto směru za neměnné (tzn. poměry se v čase nevyvíjí), lze situaci interpretovat následovně: část studentů usilujících o práci v komerční sféře nakonec našla uplatnění ve sféře státní. Lze pozorovat jasně patrný vliv pohlaví na pracovní zaměření (výstup D.5.7), přičemž v kartografických odvětvích se více uplatňují ženy, a naopak v oboru programování převažují muži.

Katederní sada mobilit (KGI_STAZ)

Ze statistické analýzy vyplývá, že v 70 % záznamů se jedná o první a současně jediný výjezd účastníka na mobilitu (výstup D.6.2). Výrazně převažujícím programem mobility je program Erasmus. Nejvíce výjezdů na mobilitu se konalo v akademickém roce 2014/2015 (výstup D.6.3), kdy na katedře probíhaly kromě tradičních programů CEEPUS a ERASMUS také výjezdy v rámci realizovaných projektů *NeoCartoLink*, *GeoInfoNet*, *InDog* a dalších (více informací o projektech je k dispozici na webových stránkách Katedry geoinformatiky UP v sekci *Projekty*). Nejčastěji navštěvovanými zeměmi za celé období jsou Polsko s celkovým počtem 34 výjezdů, Norsko s počtem 31 výjezdů a Španělsko s celkem 26 výjezdy (výstup D.6.5).

Katederní sada uchazečů (KGI_UCHAZEC)

Statistické zpracování sady potvrdilo zjištění z dotazníkové sady uchazečů, a sice že nejzastoupenějším zdrojem povědomí o studiu je internet (výstup D.7.3). Lze pozorovat i trendy ve vývoji podílů informačních zdrojů, zejména vzestup veletrhu Gaudeamus a informací od rodiče (výstup D.7.4).

7 VÝBĚR TÉMAT PRO ANALÝZY

Po přípravné fázi a základním nastudování výchozích materiálů bylo přikročeno k analýze dat. Pro tuto práci bylo nezbytné stanovit analytickou metodu hledání a ověřování hypotéz. Množství informací, které lze z dat vyčíst, je nepřehledné. Tato skutečnost je zapříčiněna povahou dat, která kromě obsáhlé atributové složky obsahují i složku prostorovou a časovou. Vzhledem k tomu vzniká velmi vysoké množství kombinací témat k analýze. Proto byla zvolena metoda konfirmačního přístupu úzkého výběru předpokladů, nežli metoda exploračního bádání a hledání vazeb napříč daty. Cílem této fáze bylo sestavit předpoklady, jejichž ověření by mohlo přinést potvrzení všeobecně míněných tvrzení nebo zajímavá zjištění.

7.1 Tematické okruhy

Výchozí materiály obsahovaly typy atributů, které bylo možné dále kategorizovat do několika tematických okruhů a témat. Zároveň se napříč daty mnohé atributy překrývaly. Jako další krok bylo zvoleno abstrahování atributů do podoby zobecněných tematických okruhů, které umožnilo na atributy nahlížet z obecnější perspektivy a snáze tak získat představu o budoucích předpokladech.

Bylo stanoveno celkem 15 témat (pohlaví, trvalé bydliště, střední škola, povědomí, úspěšnost atd.) spadajících do pěti kategorií (demografie, student, bydlení, studium, práce). V takto zobecněném a redukováném přehledu byly dále stanoveny předpoklady.

7.2 Matice témat

Vhodným způsobem vizualizace tematických kombinací je křížová matice. Do obou os matice bylo ve shodném pořadí zaneseno všech 15 definovaných témat a témata kombinována metodou „každý s každým“. Vzniklo tak celkem 105 unikátních kombinací. Témata byla záměrně zobecněna s cílem ponechání prostoru pro variabilitu interpretací jejich kombinací. Například kombinaci témat *místo výkonu praxe* a *místo pracoviště* lze interpretovat jako „Absolventi často pracují tam, kde jako studenti vykonávali praxi“, ale i jako „Kdo vykovává praxi dále od místa trvalého bydliště, je ochoten se za prací stěhovat dále“.

V první fázi byla každá z kombinací autorem formulována jako předpoklad (např. úspěšnost studenta souvisí se způsobem bydlení během studia, pohlaví souvisí s pracovním zaměřením) a subjektivně posouzena smysluplnost kombinace a zajímavost a potenciální přínos ověření předpokladu. Na základě těchto kritérií byl stanoven užší výběr předpokladů vhodných pro analýzu.

7.3 Dotazníkové šetření

S cílem zajištění objektivnosti výběru předpokladů bylo sestaveno online dotazníkové šetření. Respondenty šetření bylo sedm akademických pracovníků Katedry geoinformatiky UP, vybraných na základě oboru jejich specializace. Dotazník měl podobu sestavené křížové matice témat a byl distribuován jednotlivým respondentům napřímo, prostřednictvím e-mailu, v cloudovém nástroji *Google Sheets*. Respondenti byli vyzváni k výběru kombinací, jimi považovaných za zajímavé, a současně k poskytnutí komentáře a interpretaci, v jaké rovině a v jakém smyslu kombinaci rozumí. Celkem odpovědělo pět pracovníků, se zbývajícími dvěma byly následně kombinace prodiskutovány a byl harmonizován výběr témat. Vedlejším cílem dotazníku bylo zjistit,

jaké předpoklady respondenti považují za zajímavé k analyzování, třebaže k nim neexistují data. Na základě těchto zjištění lze sestavit doporučení (seznam témat) pro budoucí sběr dotazníků nejen na Katedře geoinformatiky UP. Podoba dotazníku je uvedena v příloze 2. *Dotazník distribuovaný členům Katedry geoinformatiky UP.*

7.4 Stanovení předpokladů

Křížové matice respondentů z navrácených dotazníků a autorem ohodnocená matice byly překryty a jednoduchou bodovací metodou ohodnoceny (podle počtu označení respondenty), přičemž s autorovým hodnocením bylo nakládáno jako s rovnocenným k hodnocení respondentů (tzn. mělo stejnou váhu). Kombinace témat s nejvyšším hodnocením byly po následných konzultacích vybrány pro další analyzování.

Posledním krokem fáze vymezení analýz byla exaktní formulace předpokladů. Ta vycházela z komentářů a interpretací kombinací stanovených respondenty (v případech, kdy byl komentář přítomen). Pokud komentář přítomen nebyl nebo nastala kolize dvou a více interpretací, znění předpokladu bylo stanoveno na základě autorova uvážení. Celkem bylo formulováno 12 předpokladů, které byly formulovány do podoby titulu (předpoklad) a podtitulu (upřesnění). Přesné znění všech předpokladů je uvedeno v tabulce 3.

Tab. 3 Předpoklady stanovené pro analýzu

Id	Předpoklad
1	Studenti se po absolvování vrací zpět do místa bydliště Trvalé bydliště před studiem a místo pracoviště po studiu jsou podobná
2	Zdroj povědomí o studiu závisí na trvalém bydlišti uchazeče Vzdálenější uchazeči získali povědomí o studiu jinými způsoby, než uchazeči bydlicí blíže
3	Účast na zahraniční mobilitě ovlivňuje studijní prospěch Studenti, kteří absolvovali zahraniční mobilitu, mají lepší studijní průměr
4	Čím dále uchazeč bydlí, tím pravděpodobněji nenastoupí Uchazeči se vzdálenějším trvalým bydlištěm mají vyšší tendenci k nenastoupení
5	Typ střední školy předurčuje prospěch studenta Studenti z gymnázií mají lepší studijní výsledky
6	Místo trvalého bydliště souvisí s úspěšností absolventa Mezi místem původu studenta a jeho prospěchem existuje spojitost
7	Kariérní směr je předurčen pohlavím studenta Pohlaví ovlivňuje, jakou kariérní cestou, zaměřením a sférou se absolvent vydá
8	Poměr zdrojů povědomí uchazečů o studium se vyvíjí Na vzestupu jsou neosobní zdroje jako internet, naopak klesají zdroje osobní
9	V čase se vyvíjí poměr zapsaných a nezapsaných uchazečů Přibývá uchazečů, kteří podali přihlášku, byli přijati, ale nenastoupili
10	Způsob bydlení během studia ovlivňuje prospěch studenta Studenti bydlicí určitým způsobem mají prospěch odlišný od jiných způsobů
11	Celková úspěšnost studentů se v čase vyvíjí Vývoj mediánového průměru studentů má jasně zřetelný trend vývoje
12	Způsob bydlení ovlivňuje, zda student absolvuje mobilitu Studenti, kteří nebydlí u rodičů, častěji vyjíždějí na zahraniční mobility

8 POKROČILÁ ANALÝZA

Analýzou studentů Katedry geoinformatiky UP se zabývala Šabatová (2019). Cílem této práce je navázat na doposud zjištěný soubor poznatků a rozšířit jej, nikoliv opakovat již provedené analýzy. Proto jsou v práci záměrně vynechány některé části popisující základní statistickou i prostorovou charakteristiku, a pozornost je věnována jiným, dosud neodvedeným analýzám.

Stanovením seznamu předpokladů byly splněny veškeré podmínky pro přikročení k samotné analýze dat. Předpoklady byly analyzovány, hodnoceny a diskutovány nezávisle na sobě. Ověření každého předpokladu vyžadovalo na míru sestavený postup analýzy. Obecným základem postupu analýzy každého předpokladu však byla úprava základových dat, základní statistická analýza, pokročilá analýza, prostorová analýza, interpretace, vyhodnocení a diskuse výsledků a na závěr tvorba grafických výstupů.

8.1 Analýza předpokladů

Prvním krokem v ověření téměř každého předpokladu byla úprava základových datových sad na míru konkrétní analýze. Typickými operacemi bylo očištění nerelevantních zbytných atributů, agregace, transformace a výpočet nových atributů, a v neposlední řadě vzájemné propojování datových sad. Klíčem pro propojení bylo ve většině případů kombinace *jméno + příjmení + akademický rok*. Pro zajímavost, v akademickém roce 2018/2019 by hrozilo, že by tato kombinace nebyla unikátní, protože v předchozím akademickém roce absolvoval bakalářské studium student stejného jména a příjmení, jako další student, který nastoupil do navazujícího studia v roce 2018/2019, nicméně původní student z bakalářského studia studium v termínu nedokončil a do navazujícího studia nenastoupil. Kombinace těchto atributů se proto pro účely této práce ukázala jako dostatečně unikátní pro použití jako klíč, a současně bylo možné tyto atributy získat pro většinu záznamů. Úprava probíhala v prostředí *RapidMiner Studio* formou tvorby vývojových modelů vycházejících ze základových datových sad.

Následovala tvorba základních statistických vizualizací nově vytvořených atributů, rovněž v prostředí *RapidMiner Studio*. Tímto způsobem byla zjištěna první fakta ohledně daného analyzovaného předpokladu, která dále stanovovala směr, jakým se bude analýza ubírat. Následně byly vytvořeny pokročilejší grafy, diagramy a infografiky. U prostorově založených předpokladů byly analýzy prováděny v prostředí *ArcGIS for Desktop*. Pro toto prostředí byla data připravována v *RapidMiner Studio* a analýza prováděna prostřednictvím obecně známých a běžně používaných prostorových operací.

Veškeré grafické výstupy odkazované v následujících odstavcích lze dohledat v hlavní příloze práce – webové prezentaci, v sekci *Pokročilá analýza*. Odkaz na webovou prezentaci je uveden v kapitole 10.4 *Webová prezentace*.

Předpoklad 1: Studenti se po absolvování vrací zpět do místa bydliště

Na základě propojení dat STAG_STUDENT a DOT_ABSOLVENT byla vytvořena výchozí datová sada, čítající celkem 77 záznamů absolventů. Elementárními atributy byly souřadnice trvalého bydliště studenta během studia a souřadnice místa pracoviště po absolvování studia.

Nově vytvořená, propojená data byla zpracována v prostředí GIS a byly vypočítány vzdálenosti přímoú linií mezi trvalým bydlištěm a Katedrou geoinformatiky UP, Katedrou geoinformatiky UP a místem pracoviště a trvalým bydlištěm a místem pracoviště. Pro výpočet linií a jejich délky byl použit nástroj *XY to Lines*. Pro vizualizaci v mapě byly odstraněny záznamy obsahující vzdálenosti větší než 300 km, protože by výrazně klesla vypovídající schopnost histogramu a grafů z důvodu nízké podrobnosti v nižších řádech vzdáleností. Tato data navíc byla poměrně raritní.

Předpoklad 2: Zdroj povědomí o studiu závisí na trvalém bydlišti uchazeče

Vstupními datovými sadami byly KGI_UCHAZEC a STAG_UCHAZEC. Atributy vstupujícími do následných analýz byly údaje o 13 různých zdrojích povědomí, kódované do binární podoby ano/ne, získané z KGI_UCHAZEC, a poloha trvalého bydliště uchazeče, získaná ze STAG_UCHAZEC. Tímto způsobem byla vytvořena výchozí datová sada čítající přesně 100 záznamů.

Prvním krokem bylo získání vzdáleností mezi trvalým bydlištěm uchazeče a Katedrou geoinformatiky UP. Vzdálenosti byly vypočítány v prostředí GIS, a to na základě přímé vzdálenosti. Vzdálenosti byly následně statisticky zpracovány. Důležitými kroky v přípravě dat byla zejména jejich agregace za jednotlivé zdroje povědomí a stanovení generalizované klasifikace.

Předpoklad 3: Účast na zahraniční mobilitě ovlivňuje studijní prospěch

Výchozí datová sada pro analýzu vznikla spojením dvou základních datových sad, STAG_STUDENT a KGI_STAZ. Dalšími příhodnými sadami byly sada DOT_STUDENT a sada DOT_ABSOLVENT, které rovněž obsahují data o zahraničních mobilitách Katedry geoinformatiky UP. Sada KGI_STAZ ale po propojení nabízela získání reprezentativnějšího a četnějšího vzorku, dotazníkové sady tedy nakonec použity nebyly.

Vytvořená sada byla metodami agregace analyzována. Z důvodu malého množství vzorků byl místo celkového studijního průměru jako vhodnější ukazatel zvolen medián. V případě použití průměru by mohlo dojít k ovlivnění výsledků vlivem extrémních údajů.

Předpoklad 4: Čím dále uchazeč bydlí, tím pravděpodobněji nenastoupí

Výchozí datová sada vznikla propojením záznamů sady STAG_UCHAZEC (informace o počtu a zápisu) a sady STAG_STUDENT (faktické nastoupení ke studiu – dokončení prvního ročníku). Vytvořená sada čítala 2 061 uchazečů.

Stěžejním krokem bylo stanovení momentu faktického nástupu uchazeče ke studiu, neboť moment nastoupení (tj. skutečné „usednutí uchazeče do lavice“) nelze z dostupných dat zjistit. Nejvíce se tomuto údaji pravděpodobně blíží údaj o zápisu ke studiu, přestože po zápisu určitá mortalita (úbytek zapsaných uchazečů) nepochybně nastává. Z toho důvodu byl jako moment nastoupení uchazeče zvolen moment zápisu ke studiu.

Vzdálenost bydliště uchazeče od KGI byla stanovena na základě síťové analýzy. Pro tvorbu sítě byla použita data z ArcČR[®] 500, liniová vrstva silnic. Silnicím byla na základě jejich kategorie přiřazena maximální rychlost a vypočtena časová nákladovost na překonání trasy. Jako vhodná jednotka byly zvoleny minuty. Obdobným způsobem byly vymezeny i spádové zóny měst s vysokoškolskou výukou GIS (výstup A.4.5) a dojezdové zóny na Katedru geoinformatiky UP (výstup A.4.6).

Města s vysokoškolskou výukou GIS byla stanovena na základě nabídky studijních oborů uvedených na webovém portálu *VysokéŠkoly.cz* k 31. březnu 2022. Vzhledem k oborovému přesahu byly zvoleny pouze školy nabízející studijní program obsahující klíčové slovo „geoinformatika“ v bakalářském nebo magisterském typu studia.

Na závěr bylo v prostředí *R Studio* provedeno statistické ověření výsledků prostřednictvím logistické regresní analýzy, na jehož základě bylo korigováno vyhodnocení předpokladu.

Předpoklad 5: Typ střední školy předurčuje prospěch studenta

Výchozí datová sada vznikla propojením univerzitní sady studentů *STAG_STUDENT*, obsahující informace o prospěchu studentů, a externí sady školských zařízení *MSMT_SKOLY* obsahující informace nezbytné ke stanovení typu střední školy a informace o kapacitě školy. Sady byly propojeny prostřednictvím IZO identifikátoru. Několik záznamů ze sady *STAG_STUDENT* se z důvodu chybějícího IZO propojit nepodařilo, jednalo se však o jednotky případů.

Byly stanoveny ukazatele prospěchu studenta (absolvoval, absolvoval s vyznamenáním, neabsolvoval, studijní průměr). Data studentů byla agregována na střední školy a byly zjištěny celkové výsledky škol o způsobu ukončení studia studentů a celkovém průměru studentů. V průběhu agregace docházelo ke značným ztrátám záznamů z důvodu neúplnosti dat o způsobu ukončení studia.

Předpoklad 6: Místo trvalého bydliště souvisí s úspěšností studenta

Použitou sadou byla sada *STAG_STUDENT*. Dodatečně byly z prostorové sady *ArcČR® 500* zjištěny informace o velikosti obce (počtu obyvatel), ze které studenti pochází. Vhodnými daty pro podrobnější analýzu by byla sada obsahující informace o kariéře absolventů po studiu. Dostupná sada *DOT_ABSOLVENT* ale neposkytuje relevantní atributy.

Jako první krok byl stanoven význam „místa bydliště“, který lze chápat ve více rovinách. Pod místem bydliště se rozumí a) velikost obce původu studenta, b) trvalé bydliště odpovídající přechodnému bydlišti (lze částečně chápat i jako bydliště u rodičů), c) vzdálenost trvalého bydliště od katedry, d) administrativní či jinou územní jednotku původu studenta (kraj apod.).

Předpoklad 7: Kariérní směr je předurčen pohlavím absolventa

Použitými daty jsou datová sada *DOT_ABSOLVENT*. Data posbírala Šabatová (2019) v rámci bakalářské práce prostřednictvím dotazníku. Sada absolventů je vedle *DOT_UCHAZEC* a *DOT_STUDENT* jednou ze tří datových sad, posbíraných v rámci práce. Sada čítá celkem 173 plnohodnotných záznamů.

Data byla v prostředí *RapidMiner Studio* očištěna o zbytné atributy a agregována podle pohlaví a pracovního zaměření a sféry zaměření. Absolutní hodnoty byly přepočteny na procentuální zastoupení v jednotlivých skupinách, čímž byl eliminován „efekt pohlaví“ (poměr mužských a ženských absolventů je 70,1 % muži a 29,9 % ženy).

Předpoklad 8: Poměr zdrojů povědomí uchazečů o studium se v čase vyvíjí

Použitá data pocházejí z interní katederní datové sady sbírané v rámci vědeckého výzkumu. Obsahují odpovědi uchazečů o studiu týkající se zdrojů povědomí. Protože se počty respondentů v letech značně mění, data byla přepočtena na relativní procentuální podíly a nad nimi byla provedena analýza.

Předpoklad 9: V čase se vyvíjí poměr zapsaných a nezapsaných uchazečů

Použity byly univerzitní datové sady uchazečů a studentů obsahující informace o výsledku zápisu uchazečů o studium. Z nich byli vyfiltrováni pouze uchazeči o 1. ročník bakalářského studia. Data byla agregována na akademické roky.

Protože byla existující sada uchazečů nekompletní, obsahující množství chybějících záznamů, byla propojena se sadou studentů, čímž byl zjištěn skutečný nástup ke studiu. Atribut o nástupu byl poté zpětně doplněn do výchozí datové sady.

Předpoklad 10: Způsob bydlení během studia ovlivňuje prospěch studenta

Použitými daty byla univerzitní datová sada obsahující informace o studijním prospěchu studentů, a dotazníkové datové sady studentů a absolventů, obsahující odpovědi týkající se způsobu bydlení v Olomouci během studií. Datová sada studentů čítala celkem 75 záznamů, datová sada absolventů celkem 173 absolventů.

Nejprve byla nachystána kombinovaná datová sada studentů a absolventů, na kterou byla napojena informace o prospěchu. Kvůli neuvedeným identifikačním údajům v dotaznících docházelo při zpracování ke značným ztrátám záznamů. Tyto údaje byly použity jako klíč pro spojení. Podařilo se napojit 136 z celkem 248 záznamů. Výsledek byl nakonec ověřen statisticky v prostředí *R Studio* metodou ANOVA.

Předpoklad 11: Celková úspěšnost studentů se v čase vyvíjí

Pro analýzu byla použita data z univerzitní datové sady studentů STAG_STUDENT, obsahující informace o prospěchu studentů. Jako vhodný ukazatel úspěšnosti studenta byl zvolen celkový průměr za celé studium, a to ve stavu na konci akademického roku, ve kterém student studium ukončil (úspěšně či neúspěšně).

Prvním krokem bylo odfiltrování záznamů studentů, kteří doposud studium nedokončili a pokračují do dalších ročníků. Dále byly odfiltrovány záznamy studentů doktorského studia a uchazečů o rigorózní řízení, a záznamy s nevalidním či zcela chybějícím průměrem. Výchozí sada čítala 984 záznamů. Data byla agregována na akademické roky a znázorněna do grafu.

Předpoklad 12: Způsob bydlení ovlivňuje, zda student absolvuje mobilitu

Použitými daty byly sady studentů a absolventů z dotazníku Šabatové (2019). Datové sady obsahují otázky ohledně způsobu bydlení během studií, a otázky, zda student vycestoval na mobilitu. K použití se rovněž nabízela katederní datová sada mobilit, která ale neposkytuje žádnou přidanou hodnotu či větší počet záznamů oproti sadám dotazníkovým.

Prvním krokem bylo spojení datových sad studentů a absolventů do jedné výchozí datové sady. Byl sjednocen datový model a harmonizovány hodnoty atributů. Výchozí datová sada čítala 136 záznamů. Data byla agregována na jednotlivé způsoby bydlení a byly vytvořeny vizualizace.

8.2 Způsob vyhodnocení

Bylo zpracováno celkem 12 analýz předpokladů. Součástí analýz bylo i vyhodnocení předpokladu ve smyslu potvrzení či vyvrácení. Optimálním způsobem hodnocení by bylo ověření prostřednictvím testování statistických hypotéz. Vzhledem k charakteru dat byl autorem tento způsob využit pouze v omezeném rozsahu a aplikován pouze na vybrané předpoklady. Bylo přistoupeno k hodnocení na základě individuálního kvalitativního

a vizuálního posouzení výstupů analýz, které ve většině případů poskytuje dostatečný vědomostní základ pro stanovení závěru.

Pro hodnocení byla stanovena čtyřstupňová škála: vyvráceno, spíše vyvráceno, spíše potvrzeno, potvrzeno. Hlavními kritérii hodnocení byla zjištění získaná z provedených analýz. Neopomenutelným kritériem však zůstala i kvalita výchozích dat, vytvořených syntézou datových sad na počátku analýzy. Ta v některých případech neposkytovala dostatečné množství záznamů pro jednoznačné potvrzení nebo vyvrácení předpokladu a snižovala tak vypovídací hodnotu závěru. Fakt, že výchozí data nedisponovala požadovanou kvalitou či uspokojivým množstvím záznamů, byla rovněž argumentem k upuštění od exaktního statistického vyhodnocení, které, vzhledem k těmto skutečnostem, nebylo považováno za dostatečně účelné.

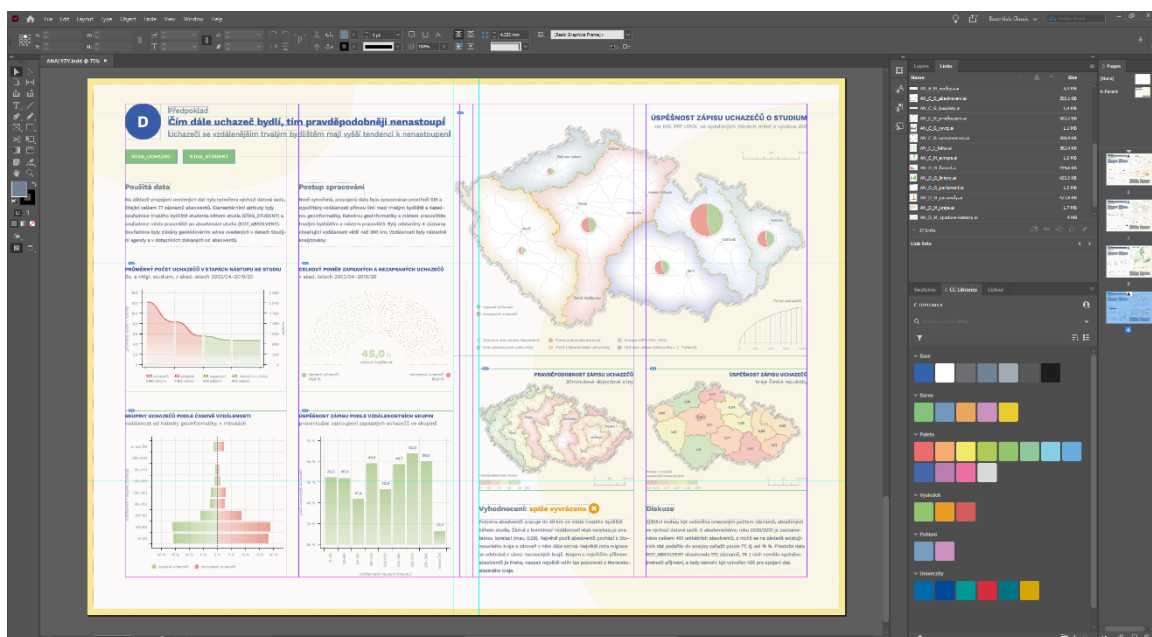
8.3 Analytické infografiky

Výstupem analýzy každého předpokladu je přehledová analytická infografika. Infografiky jsou pojaty komplexně a prezentují celý průběh analýzy od použitých dat po vyhodnocení předpokladu a diskusi výsledků.

Tvorba probíhala obdobným způsobem jako tvorba datových statistických infografik, tzn. jsou vyvedeny v jednotném grafickém provedení. Kompozice je logicky uspořádána od základních statistických grafů a vizualizací, následují pokročilejší vizualizace, v případě prostorových analýz hlavní mapový výstup a doplňující mapové výstupy.

Primární surové grafické výstupy byly získány ze softwarů *RapidMiner Studio* a *ArcGIS for Desktop* a do výsledné podoby upraveny v *Adobe Illustrator*. Následně byly v nativním formátu AI načítány do *Adobe InDesign*, kde byla tvořena výsledná kompozice (obr. 7).

Infografiky prvních šesti předpokladů jsou zpracovány ve formátu A2, zbývajících šest infografik ve formátu A4. Stejně jako u datových infografik splňují všechny náležitosti pro získání plnohodnotného tiskového výstupu (rozlíšení 300 dpi, barevný model CMYK, minimální typografické požadavky na velikost písma atd.).



Obr. 7 Tvorba kompozice analytických infografik v Adobe InDesign.

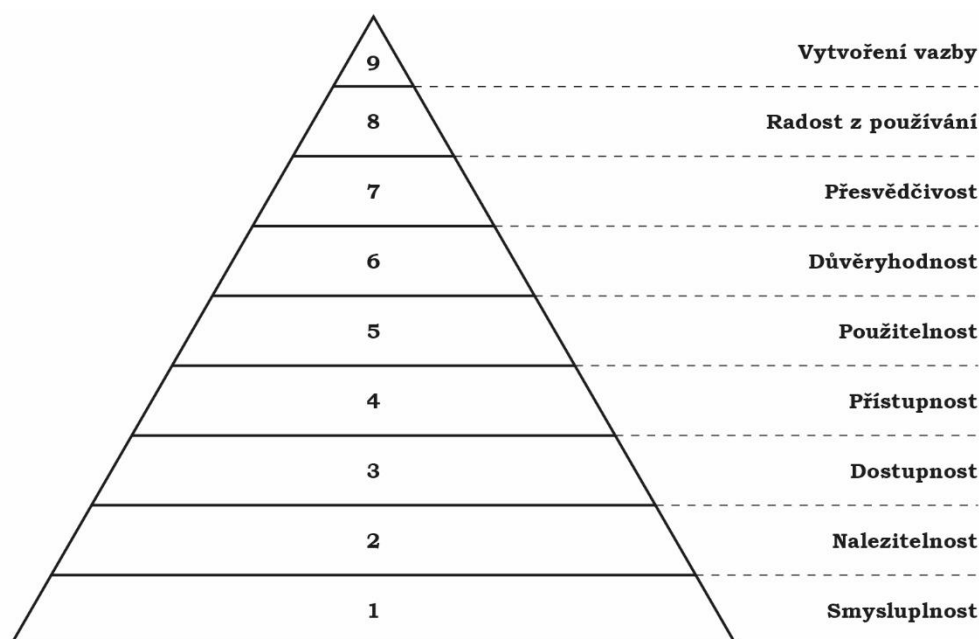
9 WEBOVÁ VIZUALIZACE

Analytická část byla odvedena, v této další fázi je cílem zjištěné poznatky prezentovat. Jako vhodný způsob prezentace zjištěných výsledků byla zvolena webová vizualizace. Webové stránky byly zvoleny zejména pro svou snadnou naležitelnost a dostupnost. Dalším argumentem volby tohoto média jsou skutečnosti vyplývající z analytické části, a sice, že nejčastěji uváděným zdrojem informací o studiu je právě internet.

Tvorba atraktivního a uživatelsky orientovaného webu není jednoduchou záležitostí. Je potřeba k ní přistupovat systematicky tak, aby nebyla vynechána žádná její důležitá fáze. V opačném případě hrozí, že sestavený web se bude např. míjet účelem, nebo nebude splňovat uživatelská očekávání. Proto byl postup výroby webu rozpracován do několika dílčích fází. Nejprve byly definovány požadavky, následně byl web navržen z hlediska prezentovaných informací i grafického zpracování, a nakonec technicky implementován a optimalizován.

9.1 Definování požadavků

Jednou z prvních fází výroby jakéhokoliv webu je stanovení požadavků na výsledný produkt. Často používanou metodou stanovování cílů reflektujících potřeby uživatelů je *Maslowova pyramida* (obr. 8). Metoda vychází z obecné Maslowovy pyramidy fyzikálních potřeb. Jedná se o logicky a hierarchicky uspořádaný model potřeb návštěvníka webu, který odráží různé úrovně webdesignu (Řezáč, 2016). Při vymezení požadavků na web bylo využito právě tohoto konceptu, tj. aby výsledný web splňoval podmínky všech pater pyramidy. Nejprve byl stanoven účel webu a jeho cílová skupina, a následně stanoveny technické požadavky vycházející z pater Maslowovy pyramidy webdesignu.



Obr. 8 Maslowova pyramida webdesignu (zdroj: Jan Řezáč; grafická úprava: autor).

9.1.1 Účel webu a cílová skupina

Hlavním cílem sestavovaného webu je prezentace výsledků analytické části práce. Web musí obsahovat všechny výstupy analýz, a to včetně všech náležitostí. Musí být jasně zřetelné, jakým způsobem bylo výsledku dosaženo a co je závěrem analýzy.

Vedlejšími cíli je sestavení webu v takové podobě, aby mohl být použit pro propagační účely, a to jak studia na Katedře geoinformatiky UP, tak jako demonstrace možností uplatnění prostorových i neprostorových analýz nad daty institucí většího měřítka (např. fakulty, univerzity). V datech o studentech se velmi pravděpodobně ukrývají informace skýtající například možnosti uplatnění v oblastech marketingu (optimalizace cílení reklamy, segmentace studentů apod.).

Stanovení cílové skupiny bylo jednou z největších výzev práce. Výchozí data či výsledky analýz nejsou svým charakterem nijak specifická pro vytyčení směru. Za cílovou skupinu, která by mohla z výsledného webu těžit nejvíce, jsou však považováni uchazeči o studium. Těm může web sloužit jako demonstrace využití prostorové analýzy, nebo jako motivace do studia. Sekundární cílová skupina je širší, obecněji zaměřená, a jedná se o všechny uživatele se zájmem dozvědět se o Katedře geoinformatiky UP a jejich studentech nové zajímavosti nebo potvrzení všeobecných předpokladů.

9.1.2 Technické požadavky

Z hlediska nejnižšího patra pyramidy (naležitelnost) bude web umístěn na snadno dohledatelné doméně, v ideálním případě na oficiální doméně katedry. Web bude splňovat základní požadavky na SEO (vyplněny klíčové meta tagy a klíčová slova, svižné načítání, obrázky mají textovou alternativu atd.). V případě otázky dostupnosti bude web vhodně strukturován pomocí sémantických značek. Vhodně strukturován bude i zdrojový kód webu, a to s ohledem na oddělení obsahové, strukturní i prezentační složky výsledného produktu. Přístupnost bude zajištěna chodem ve všech typech současných moderních prohlížečů a bezproblémovým zobrazením na různých typech zařízení (multiplatformní, responzivní). Důvěryhodnost webu bude podpořena moderním grafickým zpracováním, provedeným ve vizuálním stylu odpovídajícím vizuálnímu stylu katedry. Cílů vrchních pater pyramidy bude dosaženo za pomoci interaktivních dynamických prvků.

9.2 Návrh webu

Návrh webu je velmi podstatnou a neoddiskutovatelnou fází vývoje. Slouží jako prostředník mezi vytyčenými cíli a výsledným implementovaným produktem. Návrh byl zpracován na základě stanovených uživatelských i technických požadavků.

9.2.1 Informační architektura

Objem dat prezentovaný na webu je sice vysoký, nepřesahuje ale únosnou míru pro umístění veškerého obsahu na jednu stranu. Výhodou tohoto přístupu je jednorázové načtení veškerých dat (narozdíl od vícestránkové prezentace) a přímá dějová linie prezentovaných informací. Stránka je v logickém sledu členěna do na sebe navazujících sekcí, přičemž je postupně budována znalostní základna uživatele. Na úvod je pro zasazení webu do kontextu uvedeno stručné představení Katedry geoinformatiky UP, dále prezentace postupně prostupuje stále specifičtějšími sekcemi.

9.2.2 Drátěný model

Po sestavení základní struktury v podobě sekcí byl sestaven i jejich obsah, a to v podobě *low-fidelity* drátěného modelu. Drátěný model slouží především k rozvržení obsahu na stránce, stanovení kompozice a určení vizuální priority prvků. V žádném případě se nejedná o výsledný grafický návrh, se kterým se leckdy zaměňuje. Drátěný model byl sestaven s cílem urychlit tvorbu grafického návrhu.

9.2.3 Grafické zpracování

Samotnému grafickému zpracování předcházela rešerše grafických prvků, barevných palet a dalších vizuálních inspirací. Tato rešerše byla zpracována ve formě tzv. *mood boardu*. Jedná se v podstatě o koláž reprezentující pocit a emoce, které má sestavovaný produkt vyvolávat. Z dříve definovaných požadavků na web vyplynula podmínka využití barevné palety shodné s vizuálním stylem KGI.

Grafické zpracování probíhalo v editoru Adobe XD. Bylo využito standardního rozvržení založeném na 12 vertikálních sloupcích a rozlišení o šíři 1920 px. Barevná paleta byla sestavena autorem, ikony použité v designu pochází z volně dostupné sady ikon *Tabler Icons* (www.tablericons.com), distribuované pod licencí MIT.

9.3 Implementace

Poslední částí vývoje webu byla samotná technická implementace. V úvodu fáze bylo sestaveno a optimalizováno prostředí, kde následně probíhalo kódování a programování webové šablony. Na závěr byl web naplněn vstupními daty, tj. výsledky analytické části, texty, obrázky apod.

9.3.1 Vývojové prostředí

Centrálním prvkem prostředí byl automatizační nástroj *Gulp.js*. Skrze něj probíhalo zpracování veškerých kompilovaných jazyků (NJK, SCSS) do nativní podoby jazyků interpretovatelných prohlížečem (HTML, CSS). Pro každý z kompilovaných jazyků byl do *Gulp.js* doinstalován modul umožňující jeho zpracování. Dále byly doinstalovány balíčky umožňující minifikaci kódu (*cssnano*), spojování částí kódu (*gulp-concat*), přejmenovávání souborů (*gulp-rename*) a další. Nástroj byl nastaven na automatické sledování změn v souborech a jejich okamžitou kompilaci.

Dalším automatizačním prvkem vývojového prostředí byl nástroj *BrowserSync*, umožňující automatické obnovení stránky prohlížeče po změně ve zdrojovém souboru. Psaný kód je tak možné ihned vizuálně ověřit v prohlížeči bez nutnosti stránku manuálně obnovovat. Přítomnost podobného nástroje je obzvláště užitečná v případech inkrementálně psaných kódů, jako například kódování webových šablon.

Jako architektura pro organizaci CSS (SCSS) kódu byla zvolena konvenční *metoda 7-1*, načítající veškeré zdrojové soubory různé úrovně abstrakce do jednoho hlavního souboru. Podobným způsobem byla nastavena i architektura *Nunjucks* šablon, tj. jeden centrální soubor načítající jednotlivé sekce. Verzování kódu bylo zajištěno technologií *Git* s repozitářem hostovaným na službě *Bitbucket*. Při vývoji byl použit správce balíčků *Node.js Package Manager*.

9.3.2 Kódování šablony

HTML struktura šablony byla tvořena prostřednictvím technologie *Nunjucks*, umožňující pracovat s HTML více jako s programovacím jazykem. Bylo využito zejména funkcionality vnoření souborů (*includes*), a cyklů v kombinaci s využitím proměnných načítaných z formátů JSON (viz kapitola 9.3.3 *Obsah a optimalizace*). Cyklů bylo využito zejména v případech opakujících se HTML struktur.

Prezentační složka šablony, tj. kaskádové styly, byly kódovány v jazyce SCSS. Kompilace do nativní podoby CSS probíhala prostřednictvím nástroje *Gulp.js*. Důležitou komponentou vývoje byla knihovna *Bootstrap* ve verzi 5.1.3 poskytující řadu předdefinovaných CSS tříd. *Bootstrap* byl využit zejména pro základní rozvržení stránky, ale i jako tzv. *utility* třídy (předdefinovaná jednoúčelová CSS pravidla). CSS pravidla byla zapisována metodikou *BEM* (Block, Element, Modifier).

V šabloně je použita i řada interaktivních prvků vyžadujících programování v jazyce JavaScript. Základními kameny těchto prvků jsou knihovny *Swiper.js* a *Lightgallery*. Knihovny byly v základním nastavení importovány do projektu a následně konfigurovány jazykem JavaScript. Kompletní seznam použitých knihoven a jejich verzí je uveden v příloze 8. *Přehled vývojových JavaScript balíčků*.

9.3.3 Obsah a optimalizace

Požadavkem vyplývajícím z definovaných požadavků bylo oddělení strukturní a obsahové složky kódu. Důvodem pro separaci těchto složek byl výrazný objem obsahových dat načítaných do HTML kódu. Takový kód by se v nativním HTML jen velmi těžko spravoval.

Prvním způsobem oddělení těchto složek je využití databáze. To by však zpomalovalo nejen samotný vývoj, ale i rychlost načtení stránky v prohlížeči (stránka by musela být sestavena na straně serveru a teprve potom odeslána uživateli). Proto byl zvolen způsob statického generování stránek za pomoci technologie *Nunjucks*. S použitím cyklů a proměnných mohla být data načítána z externích JSON souborů a zkompileována do plnohodnotné výsledné HTML podoby. Tím byla zachována svižnost vývoje, a zároveň usnadněna případná další editace dat.

K optimalizaci rychlosti načítání stránek přispěla zejména optimalizace obrázků použitých na webu. Obrázky byly jednak konvertovány do webového formátu *WebP*, poskytujícího velmi dobrý poměr kvality a souborové velikosti, a jednak byly online nástrojem *TinyPNG* komprimovány. Na webu jsou pak načítány až ve chvíli jejich faktické potřeby (tzv. *lazy loading*), a to v okamžiku, kdy prohlížeč uživatele zadá požadavek na jejich zobrazení. Vše probíhá automaticky na pozadí, bez potřeby akcí uživatele.

10 VÝSLEDKY

Odvedená práce postupovala komplexním procesem zisku informace ukryté v datech, od shromáždění surových datových sad po výslednou vizualizaci. V průběhu práce bylo zjištěno množství zajímavých skutečností, byly ověřeny některé všeobecně míněné předpoklady a byla sestavena řada grafických výstupů prezentujících dosažená zjištění.

10.1 Dostupné datové zdroje

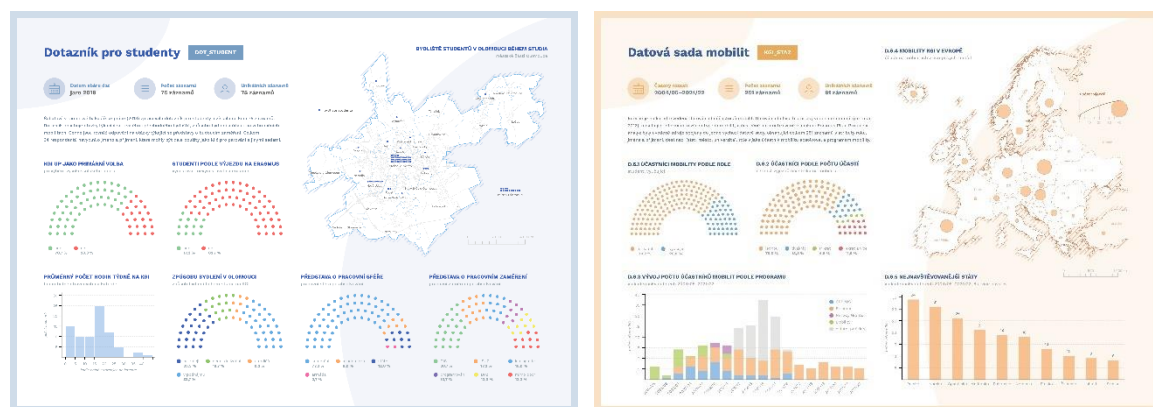
Analytické fázi předcházela rešerše existujících dat. Studovány byly akademické práce vytvořené na Katedře geoinformatiky UP analyzující studenty, uchazeče a absolventy. Dále interní datové sady vedené na KGI v rámci evidence a výzkumu, data z univerzitní Studijní agendy a další externí zdroje. U každé sady byl zjištěn atributový obsah, časový rozsah a zdroj/správce dat. Výsledkem je 12 sad spjatých s KGI a další externí sady potenciálně použitelné nejen v analýzách této práce, ale i dalším výzkumu (tabulka 4).

Tab. 4 Dostupné datové zdroje související s KGI UP

Id	Název	Zdroj/správce	Časové určení
1	Studijní agenda uchazečů	Studijní odd. PrF UP	2003/04–2020/21
2	Studijní agenda studentů	Studijní odd. PrF UP	2000/01–2020/21
3	Katederní sada uchazečů	doc. Ing. Zdena Dobešová, Ph.D.	2016/17–2019/20
4	Katederní sada mobilit	doc. RNDr. Jaroslav Burian, Ph.D.	2004/05–2021/22
5	Katederní sada praxí	RNDr. Jan Brus, Ph.D.	2017/18–2021/22
6	Katederní sada aGITací	doc. Ing. Zdena Dobešová, Ph.D.	2015/16–2018/19
7	Dotazník pro uchazeče	Bc. Marie Šabatová	jaro 2019 (sběr dat)
10	Dotazník pro studenty	Bc. Marie Šabatová	jaro 2019 (sběr dat)
11	Dotazník pro absolventy	Bc. Marie Šabatová	jaro 2019 (sběr dat)
12	Dotazníkové šetření Vichové	Mgr. Zdenka Vichová	2006 (sběr dat)

10.2 Statistické vyhodnocení

Statistickému vyhodnocení části použitých dat se věnovala Šabatová (2019). Cílem této práce bylo na tyto výsledky navázat, proto byly zpracovány pouze základní přehledy. Sady selektované pro vstup do analýz byly vyhodnoceny ve formě základních grafů a mapových výstupů. Pro každou použitou datovou sadu byla zpracována přehledová infografika statistik a základní prostorové distribuce. Celkem bylo vypracováno sedm infografik. Náhledy jsou přiloženy v příloze 3. *Náhledy statistických infografik.*



Obr. 9 Náhledy vybraných statistických infografik.

Přestože statistické vyhodnocení těchto dat nebylo cílem diplomové práce, protože bylo realizováno v bakalářské práci Šabatové (2019), na kterou tato práce navazuje, již to samotné přineslo mnohá zajímavá fakta. Proto nejzajímavější výstupy statistického zpracování byly z důvodu jejich významu a atraktivity prezentovaných informací zahrnuty do výsledků práce (tabulka 5). Zajímavosti jsou posouzeny subjektivně.

Tab. 5 Nejzajímavější fakta plynoucí ze statistického vyhodnocení

Téma	Fakt
Účast na mobilitách	Na zahraniční mobilitu vycestuje průměrně jeden z pěti studentů.
Nástup ke studiu	Ke studiu nastoupí 93,8 % uchazečů ihned po maturitě.
Úspěšnost	Studium na Katedře geoinformatiky UP úspěšně dokončí 65,2 % studentů.
Nejpočetnější ročník	Nejvyšší počet uchazečů se hlásil do akad. roku 2011/2012, a to celkem 192.
Studijní průměr	Celkový mediánový studijní průměr všech studentů je 2,26 (mezi C a D).
Vyznamenání	S vyznamenáním absolvuje 4,3 % studentů, tzn. asi jeden z dvaceti.
Poměr mužů a žen	Celkový poměr pohlaví studentů je 70,1 % mužů a 29,9 % žen.
Přijetí ke studiu	Ke studiu je každoročně v průměru přijato 69,5 % uchazečů.
Typ střední školy	Nejvíce studentů přichází z gymnázií, a to celkem 81,8 %.
Zdroj povědomí o studiu	Uchazeči jako nejčastější zdroj uvádějí internet, na druhém místě Gaudeamus.
Představa o povolání	Před nástupem ke studiu má 28,1 % uchazečů představu o budoucím povolání.
Primární volba studia	Pro 70,7 % studentů bylo studium na KGI UP primární volbou.
Způsob bydlení	Na kolejích bydlí 33,3 % studentů, v podnájmu 38,7 %. Ostatní bydlí u rodičů.
Sféra uplatnění	V komerční sféře pracuje 58,4 % absolventů.
Forma zaměstnání	89 % absolventů pracuje jako zaměstnanec. Ostatní jako OSVČ.
Nejnavštěvovanější stát	Nejnavštěvovanějším státem v rámci mobilit je Polsko s celkem 34 návštěvami.

10.3 Vyhodnocení předpokladů

Ze statisticky zpracovaných datových sad bylo stanoveno celkem 15 kandidátních tematických okruhů. Prostřednictvím online dotazníkového šetření mezi vybranými pracovníky KGI byly zjištěny nejžádanější témata a kombinace témat příhodných k analýze. Na základě výsledků dotazníku bylo formulováno 12 předpokladů. Tyto předpoklady byly následně řadou statistických i prostorových metod analyzovány, vyhodnoceny a nakonec diskutovány. Výstupem analýz je soubor poznatků v podobě 12 přehledových infografik obsahujících kompletní postup, výsledky a interpretaci analýz. Následující odstavce obsahují vyhodnocení a diskusi předpokladů. Odkazované výstupy lze dohledat v příloze 4. *Náhledy analytických infografik* nebo příloze 9. *Webová prezentace*, sekce *Pokročilá analýza* (viz kapitola 10.4 *Webová prezentace*).

10.3.1 Vyhodnocení předpokladů

Předpoklad 1: Studenti se po absolvování vrací zpět do místa bydliště

Trvalé bydliště před studiem a místo pracoviště po studiu jsou podobná

Vyhodnocení: spíše vyvráceno **X**

Polovina absolventů pracuje do 40 km od místa trvalého bydliště během studia (výstup A.1.2). Žádná z kombinací vzdáleností však nevykazuje znatelnou korelaci (max. 0,29) (výstup A.1.4). Největší podíl absolventů pochází z Olomouckého kraje a zároveň v něm dále setrvává. Největší míra migrace se odehrává v rámci moravských krajů (výstup A.1.8). Krajem s největším přílivem absolventů je Praha, naopak největší odliv lze pozorovat z Moravskoslezského kraje (výstup A.1.7).

Zjištění mohou být ovlivněna omezeným počtem záznamů obsažených ve výchozí datové sadě. K akademickému roku 2020/2021 je zaznamenáno celkem 401 unikátních absolventů, z nichž se na základě existujících dat podařilo do analýzy zařadit pouze 77, tj. asi 19 %. Přestože data DOT_ABSOLVENT obsahovala 173 záznamů, 78 z nich nemělo vyplněno jméno či příjmení, a tedy nemohl být vytvořen klíč pro spojení dat.

Předpoklad 2: Zdroj povědomí o studiu závisí na trvalém bydlišti uchazeče

Vzdálenější uchazeči získali povědomí o studiu jinými způsoby než uchazeči bydlící blíže

Vyhodnocení: spíše vyvráceno **X**

Nejčastěji uváděnými zdroji povědomí jsou Den otevřených dveří, GIS Weekend a internet. Souvislosti mezi místem trvalého bydliště a zdrojem zjištěny nebyly. Nejčastěji uváděným zdrojem v Olomouckém kraji byl internet. Naopak ve vzdálenějších krajích jako Praha či Ústecký kraj Den otevřených dveří (výstup A.2.4). Ve všech krajích jsou nejčastěji uváděny prezentační zdroje.

Analýzu lze celkově označit jako neprůkaznou, a to zejména kvůli nedostatečnému počtu záznamů v sestavené výchozí datové sadě. Nízký počet záznamů zároveň neumožnil blíže se zabývat typologií a stanovením obecně platných závěrů pro jednotlivé oblasti. Roli ve výsledku analýzy by mohla hrát i povaha výchozích dat. Pakliže jsou data sbírána na prezenčních akcích (např. *GIS Weekend*), mohl by tak být přímo ovlivněn poměr uvedených zdrojů povědomí.

Předpoklad 3: Účast na zahraniční mobilitě ovlivňuje studijní prospěch

Studenti, kteří absolvovali zahraniční mobilitu, mají lepší studijní průměr

Vyhodnocení: potvrzeno **✓**

V 15 z 16 po sobě jdoucích akademických letech měli studenti absolvující mobilitu stejný či lepší studijní průměr (výstup A.3.2), a to v průměru o 0,28 bodu. Pouze v roce 2019/2020 byl průměr vyšší. Absolventi mobilit mají oproti neabsolventům vyšší šanci úspěšného dokončení studia (výstup A.3.3), přičemž ze všech absolventů mobilit studium nedokončilo pouze 0,9 % (oproti 39,5 % neabsolventů). Dále absolventi mobilit studium častěji dokončují s vyznamenáním a méně častěji studium prodlužují.

Nelze stanovit, zda je studijní průměr kauzalitou, nebo jen koreluje s účastí na mobilitě. Možnými vysvětleními jsou vyšší studijní zdatnost absolventů mobilit, ale i například odlišná náročnost studia v zahraničí. Stejně tak příčinou může být skutečnost,

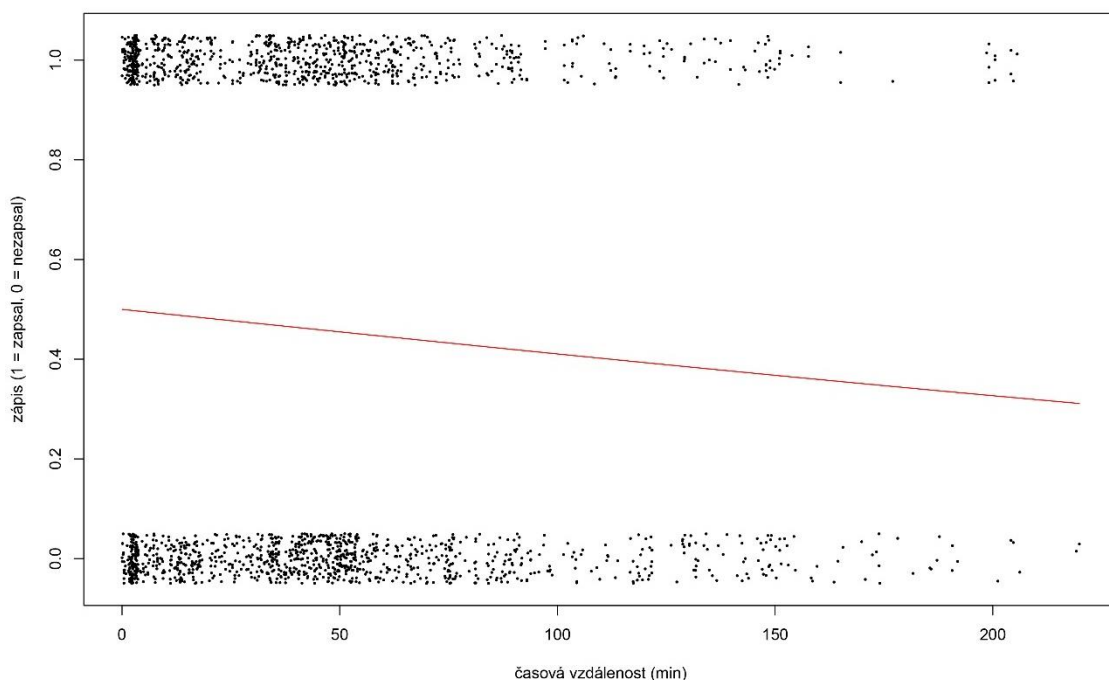
že zdatnější studenti mají potřebu více se angažovat ve studijním životě, a tedy mají vyšší tendenci vycestovat. Zjištění mohou být ovlivněna omezeným počtem záznamů obsažených ve výchozí datové sadě.

Předpoklad 4: Čím dále uchazeč bydlí, tím pravděpodobněji nenastoupí

Uchazeči se vzdálenějším trvalým bydlištěm mají vyšší tendenci k nenastoupení

Vyhodnocení: spíše potvrzeno ✓

Celková pravděpodobnost nástupu uchazeče ke studiu dosahuje 45,0 % a napříč Českou republikou se výrazně neliší (cca ± 5 %) (výstup A.4.4). Budou-li se však v mapových výstupech hledat odlišnosti i uvnitř tohoto intervalu, pak lze předpoklad považovat za mylný. Při vyhodnocení prostřednictvím lineární regrese však lze mezi úspěšností zápisu uchazeče a časovou vzdáleností pozorovat statisticky významnou závislost. Koeficient prediktoru závislosti odpovídá hodnotě 0,99, na základě čehož lze stanovit tvrzení, že s nárůstem časové vzdálenosti o 1 minutu klesá šance na zápis o 0,36 % (obr. 10). Při posouzení z hlediska vzdálenosti po silniční síti lze taktéž pozorovat statisticky významnou závislost, ačkoli mírně slabší (na 1 minutu pokles o 0,2 %). Oba ukazatele tedy mají statisticky významný vliv na zápis uchazeče ke studiu.



Obr. 10 Vyhodnocení předpokladu 4 prostřednictvím lineární regrese.

Na základě vizuálního posouzení mapových výstupů lze stanovit, že s rostoucí vzdáleností míra úspěšnosti zápisu spíše roste. Při ověření prostřednictvím regresní analýzy však byla zjištěna statisticky významná závislost vzdálenosti na úspěšnosti. Roli s největší pravděpodobností hrají odlehlé hodnoty zkreslující výsledek z pohledu celkového modelu. Zjištěné výsledky jsou platné za předpokladu, že moment faktického nástupu uchazeče ke studiu byl zvolen správně.

Předpoklad 5: Typ střední školy předurčuje prospěch studenta

Studenti z gymnázií mají lepší studijní výsledky

Vyhodnocení: spíše potvrzeno ✓

Studenti pocházející z gymnázií představují 81,8 % z celkového počtu studentů (výstup A.5.1). Mají nejvyšší podíl studentů úspěšně dokončujících studium (52 %) a nejnižší podíl neúspěšných studentů (31 %). Jedná se o jediný typ střední školy, kdy studenti dokončili studium na Katedře geoinformatiky UP s vyznamenáním (výstup A.5.4). Studenti gymnázií mají současně nejlepší celkový studijní medián (2,2). V prvních 50 studentech s nejlepším průměrem se nenachází žádný student jiného typu školy.

Přestože veškeré ukazatele hovoří ve prospěch potvrzení předpokladu, analýzu nelze považovat za zcela průkaznou z důvodu výrazně majoritního podílu studentů z gymnázií. Nizké počty z ostatních typů škol tak trpí na zkreslení způsobená úbytkem záznamů při přípravě dat a zpracováním analýzy. Jinými slovy, nezapočtením záznamů s chybějícími atributy mohlo dojít ke zkreslení výsledků minoritních typů škol.

Předpoklad 6: Místo trvalého bydliště souvisí s úspěšností studenta

Mezi místem původu studenta a jeho prospěchem existuje spojitost

Vyhodnocení: spíše vyvráceno ✗

Mezi prospěchem a místem bydliště, ve smyslu velikosti obce původu, nebylo nalezeno silnější vazby. Největší poměr úspěšných studentů pochází z obcí o velikosti 1-5 tis. obyv. (58 %), významně se však neodchyluje od jiných velikostí obcí (výstup A.6.1). Významnější rozdíl byl nalezen u porovnání trvalého a přechodného bydliště, kdy studenti bydlící na přechodném bydlišti různém od trvalého mají 64% šanci na neúspěšné absolvování studia (výstup A.6.2).

Při porovnání různých míst bydliště, na které bylo pohlíženo v různých rovinách, byly nalezeny určité odlišnosti a vzory. Při porovnání celkových průměrů územních jednotek lze pozorovat rozdíly, při vizuální analýze rozmístění absolventů ale nebylo objeveno jednoznačných prostorových vzorů či shluků. Subjektivně lze analýzu považovat za nedostatečně průkaznou.

Předpoklad 7: Kariérní směr je předurčen pohlavím absolventa

Pohlaví ovlivňuje, jakou kariérní cestou, zaměřením a sférou se absolvent vydá

Vyhodnocení: potvrzeno ✓

Z provedených analýz vyplývá zřejmá závislost mezi pohlavím a pracovním zaměřením (sférou). Cestou kartografického zaměření se vydá 20 % absolventek, ale pouze 5 % mužských absolventů (výstup A.7.1). Naproti tomu podíl programátorů můžů činí 19 %, ale pouze 4 % žen. Dále lze pozorovat o 20 % vyšší podíl absolventek pracujících ve státní sféře (výstup A.7.2).

Výsledek analýzy může být zkreslen nerovným podílem mužů a žen, který neodpovídá poměru celkovému. Zatímco celkový poměr činí 70,1 % mužů a 29,9 % žen, poměr pohlaví respondentů činí 57,2 % mužů a 42,8 % žen. Z této statistiky je zřejmé, že poměrově vůči celkovému zastoupení častěji odpovídaly ženy, čímž mohl být výsledek ovlivněn.

Předpoklad 8: Poměr zdrojů povědomí uchazečů o studium se v čase vyvíjí

Na vzestupu jsou neosobní zdroje jako internet, naopak klesají zdroje osobní

Vyhodnocení: spíše vyvráceno **X**

V diagramu (výstup A.8.1) lze pozorovat trendy ve vývoji podílů zdrojů povědomí. Na vzestupu je zejména veletrh Gaudeamus nebo informace od rodiče. Naopak na výrazném sestupu je akce GIS Weekend. Každoročně dominujícím zdrojem je internet.

Ačkoliv jsou trendy jasně zřetelné, analýzu kvůli krátkému časovému úseku nelze s jistotou stanovit za průkaznou. Dalším faktorem, který výsledky jistě ovlivnil, byla pandemie v roce 2020, čímž lze vysvětlit sestup GIS Weekendu, který se v daném akademickém roce nekonal.

Předpoklad 9: V čase se vyvíjí poměr zapsaných a nezapsaných uchazečů

Přibývá uchazečů, kteří podali přihlášku, byli přijati, ale nenastoupili

Vyhodnocení: vyvráceno **X**

Za akad. roky 2003/04–2020/21 byl zjištěn podíl zapsaných uchazečů. Lze pozorovat opačný vývoj zapsanosti uchazečů, který byl následně dokázán i lineární trendovou linií (výstup A.9.2). Podíl zapsaných uchazečů má mírně rostoucí trend.

Oproti předpokladu byl zjištěn opačný trend. Zjištěné výsledky však nelze chápat jako zvyšující se podíl uchazečů, kteří fakticky nastoupili ke studiu. Faktický nástup lze pouze aproximovat a nelze jej stanovit zcela přesně.

Předpoklad 10: Způsob bydlení během studia ovlivňuje prospěch studenta

Studenti bydlící určitým způsobem mají prospěch odlišný od jiných způsobů

Vyhodnocení: spíše vyvráceno **X**

V agregovaném prospěchu na jednotlivé způsoby bydlení byly zjištěny rozdíly. Nejlepšího studijního prospěchu dosahují studenti bydlící mimo Olomouc (medián 1,79), naopak nejhoršími způsoby bydlení z pohledu prospěchu jsou bydlení na koleji a u rodičů (medián 2,11) (výstup A.10.1). Statistické testy však významnou závislost bydlení na prospěchu nepotvrdily. Analýza variance stanovila hodnotu $p = 0,26$, na základě čehož nelze závislost považovat za statisticky významnou.

Ačkoli lze na základě vizuálního posouzení výstupů pozorovat drobné kolísání, statisticky významná závislost způsobu bydlení na prospěchu potvrzena nebyla (testováno analýzou variance a dalšími neparametrickými testy). Nesoulad ve vizuálním a statistickém posouzení je s největší pravděpodobností zapříčiněn různou velikostí sledovaných skupin a jejich velkým rozptylem hodnot.

Předpoklad 11: Celková úspěšnost studentů se v čase vyvíjí

Vývoj mediánového průměru studentů má jasně zřetelný trend vývoje

Vyhodnocení: spíše vyvráceno **X**

Z grafu (výstup A.11.1) nelze vyčíst nijak zvláště zřetelný trend vývoje. Lineární trend se sice snížil z celkové hodnoty 2,51 na 2,14, do velké míry je však ovlivněn průměrem v prvním roce měření a nízkými průměry v posledních letech měření. Po celou sledovanou dobu mediánový průměr osciluje okolo hodnoty 2,25, s výraznější odchylkou v roce 2013/14.

Budou-li pominuty akademické roky, které pravděpodobně udávají mírný vývoj trendu ve prospěch zlepšení studijního průměru (první naměřený, poslední naměřený), lze téměř po celou dobu měření sledovat podobné hodnoty. Otázkou zůstává, zda zjištěná fakta interpretovat jako konstantně se udržující úroveň studia, či přizpůsobování hodnocení úrovni studentů. Dále byla diskutována možnost vlivu aktualizace klasifikační stupnice z roku 2009/2010 odsouhlasené *Akademickým senátem PřF UP* dne 14. ledna 2009 (PřF UP, 2009) na výraznější odchylku v roce 2013/14, nicméně ta byla autorem shledána jako příliš zpožděná oproti nasazení aktualizovaného hodnocení. Roli by dále mohl hrát i nadprůměrný počet výjezdů studentů na zahraniční mobility, který v témže roce za celé sledované období (2004/05–2021/22) dosáhl svého maxima.

Předpoklad 12: Způsob bydlení ovlivňuje, zda student absolvuje mobilitu

Studenti, kteří nebydlí u rodičů, častěji vyjíždějí na zahraniční mobilitu

Vyhodnocení: spíše potvrzeno **✓**

Ze zjištěných výsledků (výstup A.12.1) vyplývá, že se během studií účastnilo zahraniční mobility 38 % respondentů bydlících na kolejích, zatímco pouze 20 % respondentů bydlících u rodičů v Olomouci. U respondentů bydlících v podnájmu je situace podobná jako u bydlících na kolejích (vycestovalo 34 %). Z výsledků je tedy patrný vliv bydlení na účast na mobilitě.

Z analýz na dotaznících bylo zjištěno, že studenti bydlící na kolejích nebo v podnájmu mají téměř dvojnásobnou pravděpodobnost účasti na mobilitě než studenti bydlící u rodičů. Analýzu ale nelze považovat za zcela průkaznou, neboť počet záznamů není považován za dostatečný. Např. u bydlení u rodičů výsledek vychází z pouhých 15 záznamů.

10.3.2 Analytické infografiky

Výstupem analýzy každého předpokladu je přehledová infografika obsahující informace o postupu, hlavní grafické výstupy analýzy, mapové výstupy, popřípadě infografiky, vyhodnocení a diskusi. Kompozice je logicky uspořádaná od představení použitých dat, přes základní statistické grafy a infografiky, po pokročilejší mapové výstupy. Infografiky jsou systematicky provázány mezi sebou a tvoří ucelenou sadu. Infografiky prvních šesti předpokladů jsou vypracovány na formát A2, ostatních šest předpokladů na formát A4. Výstupy splňují veškeré předpoklady pro plnohodnotný tisk (barevný model CMYK, rozlišení 300 dpi, minimální typografické požadavky na velikost písma atd.).

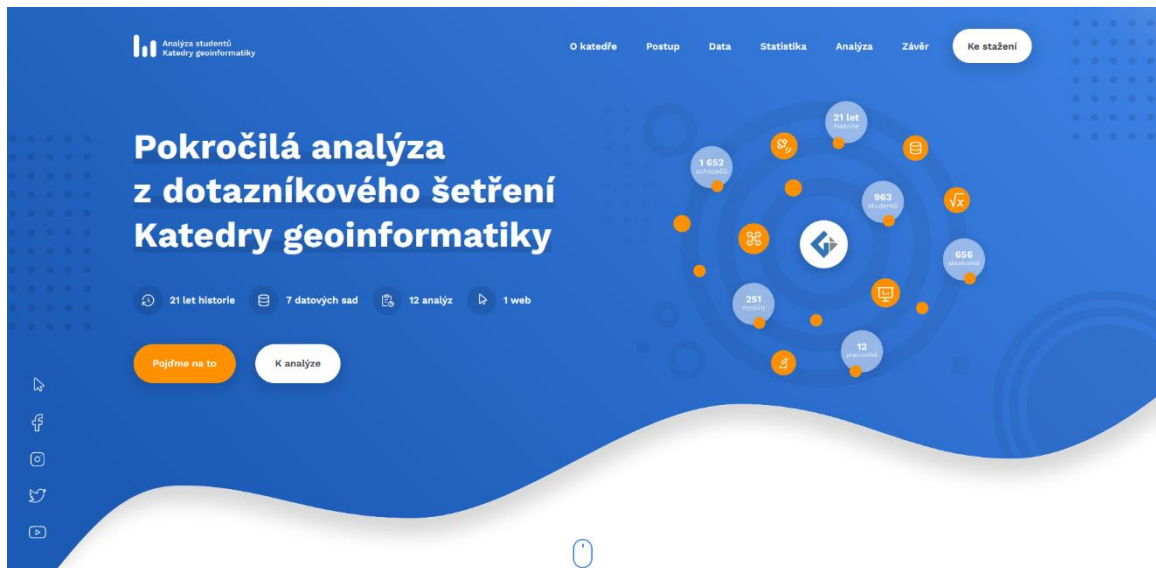
10.4 Webová prezentace

Hlavním prezentačním výstupem diplomové práce je webová vizualizace. Web přehledně dokumentuje postup a prezentuje dosažené výsledky. Stránka je strukturována do logicky na sebe navazujících sekcí postihujících náplň jednotlivých fází práce. Prochází od stručného představení Katedry geoinformatiky UP, přes postup zpracování, přehled použitých dat a výsledky provedených analýz až po diskusi a závěr.

Prezentace byla s ohledem na stanovenou cílovou skupinu (uchazeči o studium) sestavena v moderní poutavé podobě. Webu na atraktivitě přispívá množství použitých interaktivních prvků. Vizuální styl stránky je úzce spjat s vizuálním stylem Katedry geoinformatiky UP.

Výsledky analýzy předpokladů jsou prezentovány formou interaktivních karet obsahujících formulace předpokladů, jejich vyhodnocení a náhledy grafických výstupů (infografik) v podobě *lightbox* galerie. Vzhledem k měřítku některých infografik byla implementována funkcionalita přiblížení a možnost stažení.

Šablona webové stránky byla vyvinuta na míru, podle definovaných požadavků, a splňuje standardy současné webové tvorby. Web je plně funkční a zobrazitelný na všech běžných typech zařízení a ve všech moderních prohlížečích. HTML kód je validován podle *The W3C Markup Validation Service*. Web obsahuje mechanismy urychlující načtení stránky (minifikace kódu, optimalizace obrázků, lazy loading) a splňuje základní požadavky na SEO (meta tagy, textové alternativy obrázků apod.).



Obr. 12 Náhled úvodní strany webové prezentace.

<https://gis.upol.cz/analýza/>

webová prezentace je dostupná na oficiálních webových stránkách Katedry geoinformatiky Univerzity Palackého v Olomouci



10.5 Doporučení pro další sběr a práci s daty

Při kompletaci, přípravě a zpracování datových sad a analýz bylo zjištěno několik skutečností, které práci komplikovaly, zpomalovaly nebo zapříčiňovaly sníženou průkaznost analýz. S cílem usnadnění a zkvalitnění práce při dalších případných budoucích analýzách podobného charakteru byly posouzeny příčiny nedostatků a stanoveny doporučení pro další sběr dat a práci s nimi.

Zamezení redundance atributů

Mezi datovými sadami byl zjištěn značný překryv v obsažených atributech. Atributy záznamů vyskytující se v jedné sadě lze snadno dohledat v sadě jiné, a pomocí různých klíčů propojit. Ideálním klíčem k propojení datových sad se ukázala být kombinace jména, příjmení a akademického roku. Budoucí dotazníky by bylo vhodné koncipovat tak, aby obsahovaly pouze otázky, na které nelze jiným způsobem zjistit odpovědi. Tím lze docílit redukce času potřebného k vyplnění dotazníku, anebo získat prostor potřebný ke zjištění většího množství informací.

Důraz na uvedení identifikačních údajů

Ideálním klíčem pro párování záznamů jsou např. rodné číslo nebo číslo studenta. Není ale v žádném případě vhodné se na tyto údaje respondentů dotazovat. Je nutné zajistit alternativní párovací klíče z jiných dostupných atributů. Osvědčila se tvorba složeného klíče ze jména, příjmení a akademického roku, která je v naprosté většině případů unikátní. V datech však nezřídka chybí i tyto údaje. Proto by bylo vhodné respondenty více vést k jejich uvedení, například formou výzvy a zdůvodnění na konci dotazníku.

Vnitřní struktura dotazníků

Zpracovávané dotazníky byly vnitřně relativně členité a rozvětvené, čímž sice byla zvýšena informační hodnota z hlediska kvality, ale zapříčiněn vznik množství specifických skupin s nízkým počtem respondentů. S ohledem na počet záznamů v datovém souboru by bylo z hlediska kvantitativní analýzy vhodnější dotazníky členit méně. Doporučení platí zvláště pro instituce s menším počtem studentů.

Kontinuální sběr dotazníků

Jedním z přínosů s nejvyšším potenciálem pro využití poznatků v praxi jsou časové trendy. Protože dotazníkové sady ale častokrát vznikají až ve chvíli jejich faktické potřeby, datové sady nejsou kontinuální, čímž vznikají hluchá období bez existence dat. Časová řada sledovaných ukazatelů je tak nekompletní. Bylo by vhodné dotazníky sbírat průběžně a tvořit časově souvislá data, a to zejména v případě uchazečů.

Témata k zahrnutí do dotazníků

Na základě dotazníkového šetření mezi vybranými akademickými pracovníky Katedry geoinformatiky UP bylo stanoveno několik tematických okruhů pro analýzu s potenciálně zajímavým výsledkem. Pro část z nich ale nebyla dostupná relevantní nebo dostatečně podrobná data. Proto bylo shledáno vhodným doporučit do budoucích dotazníků zahrnout tato témata:

- upřesňující informace o povědomí o studiu (zahrnou kategorie sociálních sítí),
- pracovní zkušenosti nabyté během studia (např. brigády),
- čas věnovaný studiu (např. počet hodin týdně strávených studiem),
- upřesňující informace o způsobu bydlení (např. počet spolubydlících),
- upřesňující informace o zaměstnání (např. subjektivní spokojenost s prací).

11 DISKUZE

Cílem práce bylo navázat na odvedenou práci Šabatové (2019), provést pokročilou analýzu dostupných dat a vizualizovat nově nabyté, dosud nezjištěné poznatky. Postup práce byl v průběhu obou fází konfrontován s řadou překážek ohrožujících plnohodnotné dosažení vytyčených cílů. Současně bylo zjištěno několik skutečností podněcujících diskutabilnost závěrů plynoucích z výsledků.

Kvalita výchozích dat

První překážkou při zpracování analýz byla neuspokojivá kvalita některých výchozích datových sad. Část faktorů snižující kvalitu dat byla očekávána a průběhu práce potvrzena. K těmto faktorům lze zařadit zejména krátké časové řady sledovaných údajů, například údaje o uchazečích vycházejících z dotazníkových šetření. Některé stanovené předpoklady týkající se časových trendů (např. předpoklad 8, viz kapitola *8.1 Analýza předpokladů*) byly hodnoceny na základě čtyřleté řady, což lze označit za nepřesvědčivé.

Mezi faktory negativně ovlivňující kvalitu dat lze zařadit neúplnost dat ve smyslu nadměrného množství chybějících záznamů. Nabízejícím se řešením by bylo data imputovat, čímž by ale mohla být data manipulována a znehodnocena. Proto bylo zvoleno alternativní, avšak stále suboptimální řešení, a sice úplné vypuštění chybějících záznamů z analýz. To ale v mnoha případech vyústilo ve značnou redukci počtu záznamů vstupních datových sad, zejména u analýz předpokladů kombinujících více sad mezi sebou. Některé závěry tak vycházejí ze zlomku původní množiny prvků.

Použité statistické metody

Prvoplánovým záměrem pro zpracování analýz bylo použití sofistikovaných data mining metod. S pomocí takových metod by bylo z postupu možné zcela vyřadit krok vymezení předpokladů, neboť souvislosti mezi daty by byly hledány automatizovaně, a následně případně ověřeny dalšími metodami. Problematika data mining metod se ale ukázala být na osvojení časově náročnější, než bylo předpokládáno. Tyto metody by byly aplikovány na úkor dalších fází, což potenciálně ohrožovalo dosažení dalších vytyčených cílů. Proto bylo přikročeno k časově předvídatelnější konfirmační metodě analýzy předpokladů.

Přestože byly stanovené předpoklady vyhodnoceny na základě pečlivě provedených analýz poskytující dostatečný vědomostní základ, a tvrzení podpořena množstvím statistik, diagramů, map a dalších výstupů, nezpochybnitelných závěrů mohlo být dosaženo pouze za ověření prostřednictvím statistického testování. Vzhledem k povaze analyzovaných dat nebylo takové testování možné realizovat u všech předpokladů. Současně, opět vzhledem k charakteru výchozích dat ve smyslu nízkého počtu záznamů, nebylo exaktní statistické vyhodnocení považováno za dostatečně účelné. V této diplomové práci bylo statistické vyhodnocení demonstrováno pouze na vybraných předpokladech. V případě analýz rozsáhlejších datových sad je však takové vyhodnocení nevyhnutelné a zůstává tak námětem pro další práce této tematiky.

Studované zájmové území

Studium na Katedře geoinformatiky UP absolvovalo a současně i studuje řada zahraničních studentů s trvalým bydlištěm mimo území České republiky. Do předpokladů ověřujících souvislosti mezi zvolenými ukazateli a trvalým bydlištěm však tito studenti ve většině případů zahrnuti nebyli. Důvodem pro toto rozhodnutí byl jejich výrazně menšinový podíl na celkové populaci studentů (2,6 %). Rozšířením území zobrazovaného v mapových výstupech na další země by utrpěla podrobnost v hlavním

zájmovém území, což bylo, vzhledem k majoritnímu podílu studentů bydlícím v ČR, nepoměrné a nežádoucí. Současně, pokud by byla zachována konzistence sledovaných regionů napříč zeměmi, by v zahraničních regionech vznikalo mnoho regionů zcela bez dat, nebo s velmi malým vzorkem záznamů. V případě analýzy institucí větších měřítek s vyšším počtem zahraničních studentů by bylo vhodné tyto studenty zahrnout.

Zasazení výsledků do kontextu

Výsledkem práce je soubor poznatků a interpretace souvislostí v rámci studovaných datových sad. Na základě studia a analýzy použitých datových sad lze sice řadu ukazatelů kvantifikovat a získat nové informace, již ale nelze s jistotou přiřadit těmto informacím význam. Pro zcela objektivní posouzení získaných výsledků bylo by potřeba získané informace zasadit do kontextu nadřazeného celku (např. fakulta, univerzita), případně jiné srovnatelné instituce (např. další české geoinformačně orientované katedry). Teprve při vzájemném porovnání ukazatelů vyjdou najevo rozdíly a podobnosti, na jejichž základě lze informacím přiřadit užitečný význam a stanovit objektivní závěry. Protože v rámci rešerše existence srovnatelných výsledků zjištěna nebyla, bylo by potřeba sestavené analytické procesy aplikovat na odpovídající data, a tím porovnávací sadu vytvořit. Vzhledem k časové náročnosti zpracování analýz a neveřejné povaze analyzovaných dat ale nebylo možné takové srovnání provést.

Jedinými srovnatelnými výsledky jsou výsledky Vichové (2007) a výsledky Šabatové (2019). V případě vzájemného srovnání této práce a práce Vichové lze část výsledků označit na odpovídající si (např. výsledky zápisu uchazečů), ale vzhledem k rozdílným sledovaným ukazatelům a typu provedených analýz je značná část neporovnatelná. V porovnání s výsledky Šabatové dochází sice k většímu překryvu, protože však obě práce vycházejí ze stejného souboru dat, porovnání postrádá smysl.

Uplatnění výsledků

Od prvního námětu zadání práce bylo cílem formou atraktivní webové vizualizace v maximální možné míře přenést získané výsledky do praxe. Vzhledem ke skutečnosti, že nebylo možné předvídat podobu dosažených výsledků, nemohl být předem stanoven ani účel a cílová skupina zamýšleného webu. Svým všeobecným charakterem v této situaci nebyly nikterak nápomocné ani získané výsledky, které bylo nezbytné v rámci webové vizualizace objektivně prezentovat. Stanovení směru, jakým se bude webová vizualizace ubírat, se tak stalo jednou z největších výzev práce.

Po zvážení možností bylo shledáno nejvhodnějším a v praxi nejvíce uplatnitelným web orientovat na uchazeče o studium. Splnění účelu bylo podmíněno dosažením několika cílů. Prvním cílem prezentace je uchazečům blíže představit Katedru geoinformatiky UP. Uchazeč se dozví zajímavé informace nejen o KGI jako instituci, ale především informace o studentech, absolventech a dalších uchazečích. Znalost těchto faktů může sehrát důležitou roli v rozhodovacím procesu volby vysoké školy. Druhým cílem bylo demonstrovat často těžko uchopitelný proces datové analýzy a jím dosažitelné výsledky.

S ohledem na tuto cílovou skupinu je v práci kladen velký důraz na vizuální stránku prezentace výsledků, a to jak na statických grafických výstupech, tak na sestaveném webu. Prezentační stránce je autorem přikládána vysoká váha, protože právě ta může rozhodovat, zda se uživatel rozhodne svoji pozornost zobrazené stránce vůbec věnovat. V neposlední řadě může být vizuální stránka prostředkem sloužícím k dosažení nejvyšších pater Maslowovy pyramidy, což může být v konečném důsledku rozhodující konkurenční výhodou.

Nutno zdůraznit, že cílem webové vizualizace je v první řadě objektivní prezentace výsledků práce. Orientace webu na uchazeče je pouze sekundární cíl udávající obsahový směr a grafické zpracování. Záměna priorit vytyčených cílů by otevřela zcela nové možnosti prezentace. Příkladem budiž zahrnutí obsahových sekcí přibližujících studentský život na Katedře geoinformatiky UP a příklady uplatnění absolventů v České republice i zahraničí.

12 ZÁVĚR

Cílem diplomové práce bylo provést pokročilou analýzu dat a následně výsledky prezentovat ve webovém prostředí. První část práce byla zaměřena na rešerši existujících datových sad úzce spjatých s Katedrou geoinformatiky UP a jejich následnou pokročilou analýzu. Ve druhé fázi bylo cílem zjištěné poznatky atraktivní formou prezentovat, a to s ohledem na stanovenou cílovou skupinu tak, aby byl v maximální možné míře využit potenciál získaných výstupů.

Prvním krokem bylo provedení rešerše dostupných datových zdrojů potenciálně přínosných v následujících analýzách. Bylo nalezeno celkem 12 datových sad úzce souvisejících s uchazeči, studenty a absolventy KGI. Z těchto sad byly dále selektovány pouze sady relevantní k analýze. Do užšího výběru bylo vybráno sedm datových sad poskytující relevantní atributy a dostatečně kvalitní a četný vzorek. Seznámení s analyzovaným materiálem proběhlo formou základního statistického vyhodnocení a vypracováním přehledové infografiky ke každé datové sadě.

Jako postup analýzy byla zvolena konfirmační metoda definovaných předpokladů. Z obsahu datových sad bylo vymezeno 15 témat vhodných k analýze a sestavena jejich křížová matice vzájemně kombinující témata mezi sebou. Formou dotazníkového šetření mezi vybranými akademickými pracovníky Katedry geoinformatiky UP byly vybrány nejzajímavější kombinace témat. Na jejich základě bylo sestaveno 12 předpokladů k ověření. Jednotlivé předpoklady byly statistickými a prostorovými metodami analyzovány a vyhodnoceny jako potvrzené nebo vyvrácené. Výstupem analýzy každého předpokladu je přehledová infografika.

Soubor zjištěných poznatků je prezentován formou webové vizualizace. Prezentace byla implementována prostřednictvím technologií moderního vývoje webu a splňuje veškeré požadavky současného webdesignu. Účelem webu je mimo samotnou prezentaci výsledků i podpora propagační činnosti Katedry geoinformatiky UP. Web je orientován na uchazeče o studium a koncipován jako bližší představení katedry uchazečům a demonstrace uplatnění datové a prostorové analýzy.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

Adobe Creative Cloud [online]. 2022 [cit. 2022-04-15]. Dostupné z:
<https://www.adobe.com/cz/creativecloud.html>

ACKOFF, R. L. *From Data to Wisdom. Journal of applied systems analysis.* 1989.

ALTMANNOVÁ, M., 2014. *Dojíždka do vysokých škol v Brně s důrazem na studenty ze Slovenska.* Diplomová práce. Masarykova univerzita v Brně.

ArcČR® 500 [online]. ©ArcČR, ARCDATA PRAHA, ZÚ, ČSÚ, 2019. Praha: ARCDATA PRAHA, s.r.o., Zeměměřický úřad, ČSÚ, 2022. Dostupné z:
<https://www.arcdata.cz/produkty/geograficka-data/arccr-500>

BELLINGER, G., CASTRO, D., MILLS, A. *Knowledge Management – Emerging Perspectives.* 2004.

BEM Methodology. BEM.info [online]. 2022 [cit. 2022-04-15]. Dostupné z:
<https://en.bem.info/methodology/>

BUNCHENKO, D. *7-1 Sass Architecture.* Medium.com [online]. 2021 [cit. 2022-04-15]. Dostupné z:
<https://denis-bunchenko.medium.com/7-1-sass-architecture-da69da631ebe>

Cambridge Dictionary: English Dictionary, Translations & Thesaurus [online]. Cambridge University Press [cit. 2022-04-15]. Dostupné z: <https://dictionary.cambridge.org/>

COLUMBU, S., PORCU, M., PRIMERANO, I., SULIS, I., VITALE, M. P. Geography of Italian student mobility: A network analysis approach. *Socio-Economic Planning Sciences* [online]. 2021, 73 [cit. 2022-04-23]. ISSN 00380121. Dostupné z: doi:10.1016/j.seps.2020.100918

ČSÚ. *Veřejná databáze* [online]. 2022. Dostupné z: <https://vdb.czso.cz/vdbvo2/>

Dictionary by Merriam Webster: America's most trusted online dictionary [online]. [cit. 2022-04-15]. Dostupné z: <https://www.merriam-webster.com/>

DOBEŠOVÁ, Z. *Databázové systémy v GIS.* Olomouc: Univerzita Palackého, 2004. ISBN 80-244-0891-0.

DOBEŠOVÁ, Z. *Discovering association rules of information dissemination about Geoinformatics university study.* 2018.

DOBEŠOVÁ, Z., PIÑOS, J. *Using Decision Trees to Predict the Likelihood of High School Students Enrolling for University Studies.* 2019.

FIRESTONE, J. *Knowledge Management Systems and The Knowledge Warehouse. A „Strawman“.* 2000.

HAKENOVÁ, M., 2009. *Prostorová a oborová diferenciacie uchazečů o studium na PřF UP v Olomouci.* Bakalářská práce. Univerzita Palackého v Olomouci.

HAN, J., KAMBER, M., PEI, J. *Data mining: concepts and techniques.* 3rd ed. Amsterdam: Elsevier/Morgan Kaufmann, 2012. The Morgan Kaufmann series in data management systems. ISBN 978-0-12-381479-1.

HENDRICKSON, B., ROSEN, D., AUNE, R. K. An analysis of friendship networks, social connectedness, homesickness, and satisfaction levels of international students. *International Journal of Intercultural Relations* [online]. 2011, 35(3), 281-295 [cit. 2022-04-18]. ISSN 01471767. Dostupné z: doi:10.1016/j.ijintrel.2010.08.001

HEO, J., CHUNG, K., YOON, S., YUN, S. B., MA, J. W., JU, S.. Spatial-Data-Driven Student Characterization in Higher Education. In: *Proceedings of the 1st ACM SIGSPATIAL Workshop on Prediction of Human Mobility* [online]. New York, NY, USA: ACM, 2017, 2017-11-07, s. 1-4 [cit. 2022-04-23]. ISBN 9781450355018. Dostupné z: doi:10.1145/3152341.3152343

CHAUDHURI, S., DAYAL, U. *An Overview of Data Warehousing and OLAP Technology*. 1997.

INMON, W. H. *Building the Data Warehouse*. 4th ed. Indianapolis, Indiana: Wiley Publishing, 2005, 543 s. ISBN 978-0-7645-9944-6.

Internetová jazyková příručka [online]. Praha: Ústav pro jazyk český AV ČR [cit. 2022-04-15]. Dostupné z: <https://prirucka.ujc.cas.cz/>

KAMRUZZAMAN, M., HINE, J., GUNAY, B., BLAIR, N. Using GIS to visualise and evaluate student travel behaviour. *Journal of Transport Geography* [online]. 2011, 19(1), 13-32 [cit. 2022-04-18]. ISSN 09666923. Dostupné z: doi:10.1016/j.jtrangeo.2009.09.004

LACKO, L. *Databáze: datové sklady, OLAP a dolování dat s příklady v Microsoft SQL Serveru a Oracle*. Brno: Computer Press, 2003, 486 s. ISBN 80-722-6969-0.

CLARKE, G. P., LANGLEY, R. A Review of the Potential of GIS and Spatial Modelling for Planning in the New Education Market. *Environment and Planning C: Government and Policy* [online]. 1996, 14(3), 301-323 [cit. 2022-04-23]. ISSN 0263-774X. Dostupné z: doi:10.1068/c140301

LINHARTOVÁ, P., 2016. *Prostorové chování studentů Institutu geoinformatiky*. Bakalářská práce. Vysoká škola Báňská – Technická univerzita Ostrava.

LUBIENSKI, CH., DOUGHERTY, J. Mapping Educational Opportunity: Spatial Analysis and School Choices. *American Journal of Education* [online]. 2009, 115(4), 485-491 [cit. 2022-04-23]. ISSN 0195-6744. Dostupné z: doi:10.1086/599783

Ministerstvo školství, mládeže a tělovýchovy [online]. [cit. 2022-04-15]. Dostupné z: <https://www.msmt.cz/>

MŠMT: *Rejstřík škol a školských zařízení* [online]. 2022. Dostupné z: <https://rejstriky.msmt.cz/rejskol/>

NORDVIK, H., AMPONSAH, B. Gender Differences in Spatial Abilities and Spatial Activity Among University Students in an Egalitarian Educational System. *Sex Roles* 38, 1009–1023 (1998). Dostupné z: <https://doi.org/10.1023/A:1018878610405>

NORTH, M. *Data Mining for the Masses*. 2. vydání. CreateSpace Independent Publishing Platform, 2016, 252 s. ISBN 978-0615684376.

Nunjucks: A rich and powerful templating language for JavaScript. [online]. 2022 [cit. 2022-04-15]. Dostupné z: <https://mozilla.github.io/nunjucks/>

PETRÁŠ, L., 2011. *Prostorová diferenciacie studentů Univerzity Palackého v Olomouci*. Diplomová práce. Univerzita Palackého v Olomouci.

- PIKE, G.R., KUH, G.D. & MCCORMICK, A. C. An Investigation of the Contingent Relationships Between Learning Community Participation and Student Engagement. *Res High Educ* 52, 300–322 (2011). Dostupné z: <https://doi.org/10.1007/s11162-010-9192-1>
- POKORNÝ, M., 2010. *Zpracování výsledků dotazníkového šetření absolventů Přírodovědecké fakulty Univerzity Palackého v Olomouci*. Bakalářská práce. Univerzita Palackého v Olomouci.
- Zápis č. 5 ze zasedání Akademického senátu PřF UP dne 14. ledna 2009. *Přírodovědecká fakulta Univerzity Palackého v Olomouci* [online]. 2009. Dostupné z: http://old.prf.upol.cz/fileadmin/user_upload/PrF-dokumenty/AS_PrF/zapisy/zas_05-2009-01-14.pdf
- PUČAN, S., 2010. *Prostorové analýzy studentů a absolventů oboru geoinformatika na VŠB-TU Ostrava*. Bakalářská práce. Vysoká škola Báňská – Technická univerzita Ostrava.
- RapidMiner: Best Data Science & Machine Learning Platform* [online]. 2022 [cit. 2022-04-15]. Dostupné z: <https://rapidminer.com/>
- ROESSINGH, H. Variability in ESL Outcomes: The Influence of Age on Arrival and Length of Residence on Achievement in High School. *TESL Canada Journal* [online]. 2008, 26(1), 87-107 [cit. 2022-04-18]. ISSN 1925-8917. Dostupné z: doi:10.18806/tesl.v26i1.392
- ROBARDS, T. *Structuring your Sass Projects*. *ITNext* [online]. 2018 [cit. 2022-04-15]. Dostupné z: <https://itnext.io/structuring-your-sass-projects-c8d41fa55ed4>
- RUNKLER, T. A. *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. 3rd edition. Mnichov, Německo: Springer Vieweg, 2020, 161 s. ISBN 978-3-658-29778-7.
- RYKR, J., 2013. *Prostorové analýzy a vizualizace školských dat Olomouckého kraje*. Diplomová práce. Univerzita Palackého v Olomouci.
- ŘEHÁK, J, ŘEHÁKOVÁ, B. Logitové modely: analýza vlivu exogenních faktorů u kategorizovaných dat. *Sociologický časopis*. 1992.
- ŘEZÁČ, J. *Web ostrý jako břitva: návrh fungujícího webu pro webdesignery a zadavatele projektů*. Vydání druhé. Brno: House of Řezáč, 2016, 210 s. ISBN 978-80-270-0644-1.
- Sass Guidelines [online]. 2022 [cit. 2022-04-15]. Dostupné z: <https://sass-guidelin.es/>
- Sass: Syntactically Awesome Stylesheets: CSS with superpowers* [online]. 2022 [cit. 2022-04-15]. Dostupné z: <https://sass-lang.com/>
- SCHNEIDEROVÁ, N., 2017. *Dojíždka studentů ekonomických fakult za studiem v Brně*. Bakalářská práce. Masarykova univerzita v Brně.
- ŠABATOVÁ, M., 2019. *Analýza a vizualizace prostorové diferenciac studentů a absolventů Katedry geoinformatiky UP*. Bakalářské práce, Univerzita Palackého v Olomouci.
- ŠARMANOVÁ, J. *Metody analýzy dat: učební text*. Ostrava: Vysoká škola báňská – Technická univerzita Ostrava, 2012, 170 s. ISBN 978-80-248-2565-6.
- ŠULC, O. *Prognostika od A do Z*. Praha: SNTL, 1987. 210 s.
- TRULOCK, V. *Moodboards* [online]. 2008 [cit. 2022-04-15]. Dostupné z: <http://hci.ilikecake.ie/envisionment/moodboards.htm>

VÍCHOVÁ, Z., 2007. *Prostorové analýzy uchazečů, studentů a absolventů geoinformatiky na UP Olomouc*. Bakalářská práce. Univerzita Palackého v Olomouci.

VONDRÁKOVÁ, A., VOŽENÍLEK, V., RYCHTAŘÍKOVÁ, J. *Atlas vysokého školství v České republice 2001–2014*. Olomouc: Univerzita Palackého v Olomouci, Edice M.A.P.S. 2017. 115 s. ISBN 978-80-244-5139-8

VOŽENÍLEK, V. *Katedra geoinformatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci 2001–2016*. Olomouc: Univerzita Palackého v Olomouci pro Katedru geoinformatiky, 2016, 116 s. ISBN 978-80-244-5027-8.

VOŽENÍLEK, V., KAŇOK, J. a kol. *Metody tematické kartografie: vizualizace prostorových jevů*. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2011. ISBN 978-80-244-2790-4.

ZÁPADOČESKÁ UNIVERZITA V PLZNI. *Informační systém studijní agendy* [online]. Plzeň, 2022 [cit. 2022-04-15]. Dostupné z: <https://is-stag.zcu.cz/>

PŘÍLOHY

SEZNAM PŘÍLOH

Vázané přílohy

- Příloha 1 Sada přehledů atributů datových sad
- Příloha 2 Dotazník distribuovaný členům Katedry geoinformatiky UP
- Příloha 3 Náhledy statistických infografik
- Příloha 4 Náhledy analytických infografik
- Příloha 5 Náhled webové prezentace
- Příloha 6 Architektura projektu webové prezentace
- Příloha 7 Architektura kaskádových stylů webové prezentace
- Příloha 8 Přehled vývojových JavaScript balíků

Volné přílohy

- Příloha 9 Webová prezentace (www.gis.upol.cz/analyza)
- Příloha 10 Plakát (formát A2)
- Příloha 11 CD (kompletní soubor výstupů)

Popis struktury CD

```
root
|- text_prace
|   |- frycak_dp.pdf           # text práce (formát PDF)
|- export_data
|   |- jpeg                   # infografiky dat (formát JPEG, 300dpi)
|   |- pdf                     # infografiky dat (formát PDF)
|- export_analyzy
|   |- jpeg                   # infografiky analýz (formát JPEG, 300dpi)
|   |- pdf                     # infografiky analýz (formát PDF)
|- webova_prezentace/
|   |- assets/                 # zdrojové soubory: SCSS, IMG, JS
|   |- data/                   # zdrojové soubory: JSON data
|   |- html/*                  # zdrojové soubory: NJK šablony
|   |- public/*                # veřejné soubory: distribuční kód
|   |- package.json            # instalační soubor knihoven
|   |- readme.md               # dokumentace projektu
```

Poznámka: Analyzovaná data (data ze Studijní agendy UPOL a data z katederních sběrů) jsou vzhledem k jejich neveřejné povaze uloženy u vedoucí práce RNDr. Aleny Vondrákové, Ph.D., LL.M. (kontakt: alena.vondrakova@upol.cz)

Příloha 1.1 – Přehled atributů datové sady STAG_UCHAZEC

ID	Atribut	Příklad	Vysvětlení
1	PRIJMENI	NOVÁK	příjmení studenta
2	JMENO	Tomáš	křestní jméno studenta
3	OSOB_IDNO	20329	osobní identifikátor studenta
4	UCHAZEC_IDNO	25278	osobní identifikátor uchazeče
5	DATNAR	08.10.1981	datum narození studenta
6	ROKNAR	1981	rok narození studenta
7	TITUL_PRED	Bc.	titul před jménem
8	TITUL_ZA	Ph.D.	titul za jménem
9	POHLAVI	M	pohlaví studenta
10	ULICE	Ondráčkova 64	ulice a č. p. přechodného bydliště
11	OBEC	Brno	obec přechodného bydliště
12	PSC	62800	PSC přechodného bydliště
13	MISTO	Brno 28	část obce přechodného bydliště
14	STAT	Česká republika	stát přechodného bydliště
15	ULICE_TRV	Za zahradami 3	ulice a č. p. trvalého bydliště
16	OBEC_TRV	Šternberk	obec trvalého bydliště
17	PSC_TRV	78501	PSC trvalého bydliště
18	MISTO_TRV	Šternberk 1	část obce trvalého bydliště
19	STAT_TRV	Česká republika	stát trvalého bydliště
20	STAT_NAR	Česká republika	stát narození
21	MATURITA	1998	maturitní rok studenta
22	ODKUD_SKOLA	Gymnázium Šternberk	škola původu studenta
23	P_OBCANSTVI	T	české státní občanství
		C	cizí státní občanství
25	DAT_PODANI_EP	17.02.2010	datum podání elektronické přihlášky
26	FORMA	P	prezenční forma studia
		K	kombinovaná forma studia
27	TYP	B	bakalářský typ studia
		N	navazující typ studia
		D	doktorský typ studia
28	STAND_DELKA	3	standardní délka zvoleného studia
29	KOD	B1301	identifikátor studijního programu
30	PROGRAM	Geografie	studijní program
31	OBOR	Geografie-geoinformatika	studijní obor
32	BODY	65	body získané při přijímacích zkouškách
33	OZNACENI	GEINF	zkratka oboru studia
34	PORADI_BODU	66-67	celkové umístění na přijímacích zkouškách
35	POCET_PRIJATYCH	58	počet přijatých studentů do ročníku
36	KOD_ROZHODNUTI	10	přiját na základě přijímací zkoušky
		15	přiját na jiný program (obor)
		20	nepřiját pro neprospěch u přijímací zkoušky
		21	nepřiját, nedoložil doklad o absolvování SŠ
		30	nepřiját pro nedostačující kapacitu oboru
		71	přiját na VŠ bez přijímací zk.
		79	přiját na VŠ bez přijímací zk. – jiný důvod
		80	přiját na VŠ po odvolání – autoremedura
		82	nepřiját po odvol. pro neprospěch u přijímací zkoušky
		83	nepřiját po odvolání pro nedostačující kapacitu oboru
85	přiját dodatečně – doplnění kapacity		
90	neúčastnil se přijímací zkoušky		
37	DATUM_ROZHODNUTI	15.08.2011	datum rozhodnutí
39	VYSLEDEK_ZAPISU	uchazeč se zapsal	výsledek zápisu
42	DNESNI_DATUM	31.01.2019	datum exportu záznamu ze studijní agendy

Příloha 1.2 – Přehled atributů datové sady **STAG_STUDENT**

ID	Atribut	Příklad	Vysvětlení
1	OS_CISLO	R00462	osobní číslo studenta
2	STUDENTMALE	Petr Novák	jméno a příjmení studenta
3	AKAD_ROK	2000/2001	akademický rok, pro který je záznam veden
4	TITUL_PRED	Bc.	titul před jménem
5	PRIJMENI	Novák	příjmení studenta
6	JMENO	Petr	křestní jméno studenta
7	TITUL_ZA	DiS.	titul za jménem
8	POHLAVI	M	pohlaví studenta
9	ROKNAR	1980	rok narození studenta
10	ULICE	17. listopadu 50	ulice a č. p. přechodného bydliště
11	PSC_CISTE	77900	poštovní směrovací číslo přechodného bydliště
12	MESTO	Olomouc	obec přechodného bydliště
13	STAT	Slovenská republika	stát přechodného bydliště
14	ULICE_CISLO_TRV	Andělská 12	ulice a č. p. trvalého bydliště
15	OBEC_CAST_TRV	Olomouc	část obce trvalého bydliště
16	PSC_CISTE_TRV	77900	poštovní směrovací číslo trvalého bydliště
17	STAT_TRV	Slovenská republika	stát trvalého bydliště
18	IZO_SS	000601675	identifikační znak organizace, identifikátor SS
19	MISTO_SS	Palackého 50 Nový Jičín	adresa střední školy
20	NAZEV_SS	Slovanské gymnázium	název střední školy
21	STAV	N	student nestuduje
		P	student přerušil
		S	student studuje
22	PROGRAM	Geografie	program studia
23	CIS_PR	B1301	identifikátor programu studia
24	CISOBORU	1301R006	identifikátor oboru studia
25	OBOR	Geografie-geoinformatika	obor studia
26	ZKRATKA	GEOINF	zkratka oboru studia
27	ZISK_TITUL	Bc.	titul po úspěšném absolvování studia
28	STANDARD_DELKA	3	standardní délka studia
29	ROCNIK	2	studentem aktuálně studovaný ročník
30	UKONCENI	2003	předpokládaný rok ukončení studia
31	KREDITY_ZAPS	48	kredity zapsané pro aktuální akademický rok
32	KREDITY_DOS	48	kredity získané v aktuálním akademickém roce
33	KREDITY_CELK	182	celkem doposud získané kredity za studium
34	DATUM_UKONCENI	02. 10. 2002	datum ukončení studia
35	KOD_UKONCENI	1	úspěšné ukončení studia
		2	zanechání studia
		3	ukončení pro nesplnění studijních podmínek
		9	přestup na jiný program nebo školu
		99	uzavření ročníku
36	ODSTUD_DOBA_CISTA	736	počet doposud studovaných dní
37	CELKOVY_VYSLEDEK	Absolvoval	celkový výsledek studia
38	FORMA_ZK	P	forma studia
39	PRUMER_CELK	2,50	celkový studijní průměr
40	PRUMERY_ROCNI	2,5;2,5;2,48;2,65	roční studijní průměry
41	KREDITY_DOS_ROCNI	45;62;68;0	roční získané kredity
42	P_SKOLITEL	Mgr. Pavel Tuček, Ph.D.	školitel disertační práce
43	ISP	ZL	individuální studijní plán
44	PREDCH_STUDIA	ČVUT fakulta dopravní: 0	předchozí studia
45	DNESNI_DATUM	31. 01. 2019	datum exportu záznamu ze studijní agendy
46	POZNAMKA_ROCNI	2 x neobhájil	poznámka

Příloha 1.3 – Přehled atributů datové sady DOT_UHAZEČ

ID	Atribut	Příklad	Vysvětlení
1	DATUM_SUB	19.04.2019	datum odeslání dotazníku respondentem
2	TYP	Bc.	typ studia, do kterého se uchazeč hlásí
3	SS_NAZEVI	Gymnázium Šternberk	název střední školy, ze které se uchazeč hlásí
4	SS_MESTO	Šternberk	město střední školy, ze které se uchazeč hlásí
5	SS_PSC	78306	PSC školy, ze které se uchazeč hlásí
6	SS_STAT	Česká republika	stát střední školy, ze které se uchazeč hlásí
7	UCAST_GD	A	účast uchazeče na akci GIS Day
8	UCAST_GW	A	účast uchazeče na akci GIS Weekend
9	UCAST_DOD	A	účast uchazeče na dni otevřených dveří UP
10	POVED_PRED	N	povědomí o KGI z přednášky experta na SŠ
11	POVED_KAMA	N	povědomí o KGI od kamarádů a známých
12	POVED_INTE	N	povědomí o KGI z internetu
13	POVED_UCIT	N	povědomí o KGI od učitele na SŠ
14	POVED_GAUD	N	povědomí o KGI z akce Gaudeamus
15	POVED_MATE	N	povědomí o KGI z propagačních materiálů
16	NASTUP_PO	po maturitě	návaznost na předchozí studium na SŠ
17	PREDSTAVA_BUD	A	uchazeč disponuje představou o své budoucnosti
18	POHLAVI	M	pohlaví studenta
19	OBEC	Bohuňovice	obec trvalého bydliště uchazeče
20	PSC	78314	PSC trvalého bydliště uchazeče
21	STAT	Česká republika	stát trvalého bydliště uchazeče
22	VEK	19	věk uchazeče
23	JMENO	Tomáš	jméno uchazeče
25	PRIJMENI	NOVÁK	příjmení uchazeče

Příloha 1.4 – Přehled atributů datové sady DOT_STUDENT

ID	Atribut	Příklad	Vysvětlení
1	DATUM_SUB	19.04.2019	datum odeslání dotazníku respondentem
2	STUDIUM_TYP	Mgr	typ aktuálně studovaného studia
3	STUDIUM_ROCNIK	2	aktuálně studovaný ročník daného studia
4	STUDIUM_PRODLOUZENI	A	informace o prodloužení studia
5	SS_NAZEV	Gymnázium Legionářů	název střední školy studenta
6	SS_MESTO	Příbram	město střední školy studenta
7	SS_PSC	26101	PSČ střední školy studenta
8	SS_STAT	Česká republika	stát střední školy studenta
9	PRIM_VOLBA	A	geoinformatika byla primární volbou při výběru VŠ oboru
10	BYDLENI_ZPUSOB	na koleji	způsob bydlení v Olomouci
11	BYDLENI_CAST_MESTA	Nová Ulice	městská část bydlení v Olomouci
12	BYDLENI_TRVALE	N	bydliště je současně trvalé
13	BYDLENI_KOLEJ	Kolej 17. listopadu	kolej, na které student bydlí
14	DOJIZDZENI_FREKVENCE	1x týdně	frekvence dojíždění do Olomouce
15	DOJIZDZENI_ZPUSOB	vlak	způsob dojíždění do Olomouce
16	STUDIUM_HODNOCENI	4	hodnocení studia studentem (1 = ucházející; 5 = velmi přínosné)
17	STUDIUM_NAVIC	A	vzdělávání studenta nad rámec studia
18	BUD_ZAMERENI	gis	představa studenta o budoucím zaměření
19	BUD_SFERA	komerční	představa studenta o sféře
20	BUD_PREDSTAVA	A	student má představu o svém budoucím zaměstnání
21	PRAC_ZKUSENOST	A	pracovní zkušenost v průběhu studia
22	ERASMUS	A	student vycestoval alespoň na jeden ERASMUS
23	ERASMUS_POCET	1	počet účastí na ERASMU
25	LETSKOLA	A	student se zúčastnil alespoň jedné letní školy
26	LETSKOLA_POCET	1	počet účastí na letní škole
27	KATEDRA_DOPRAVA	pěšky	nejčastější způsob dopravy studenta na KGI
28	KATEDRA_DOPRAVA_CAS	15	průměrný čas strávený cestou na KGI
29	KATEDRA_STRAV_CAS	12	průměrný čas strávený týdně na KGI
30	POHLAVI	M	pohlaví studenta
31	OBEC	Konice	obec trvalého bydliště studenta
32	PSC	79852	PSČ trvalého bydliště studenta
33	STAT	Česká republika	stát trvalého bydliště studenta
34	JMENO	Tomáš	křestní jméno studenta
35	PRIJMENI	NOVÁK	příjmení studenta

Příloha 1.5 – Přehled atributů datové sady DOT_ABSOLVENT

ID	Atribut	Příklad	Vysvětlení
1	DATUM_SUB	17.12.2019	datum odeslání dotazníku respondentem
2	STUDIUM_ABSOLV	Mgr	typ nejvyššího absolvovaného studia
3	SS_NAZEV	Gymnázium Český Těšín	název střední školy
4	SS_MESTO	Český Těšín	město střední školy
5	SS_PSC	73701	PSČ střední školy
6	BYDLENI_ZPUSOB	na koleji	způsob bydlení během studia
7	BYDLENI_CAST_MESTA	Nová Ulice	část města, ve které student během studia bydlel
8	BYDLENI_TRVALE	N	uvedené bydliště bylo současně trvalé bydliště
9	BYDLENI_KOLEJ	Kolej Šmeralova	kolej, na které student během studia bydlel
10	DOJIZDENI_FREKVENCE	1x týdně	frekvence dojíždění studenta do Olomouce
11	DOJIZDENI_ZPUSOB	vlak	způsob dojíždění studenta do Olomouce
12	STUDIUM_HODNOCENI	4	hodnocení studia studentem (1 = ucházející; 5 = velmi přínosné)
13	ERASMUS	A	student vycestoval alespoň na jeden ERASMUS
14	ERASMUS_PO CET	2+	počet účastí na ERASMU
15	LETSKOLA	A	student se zúčastnil alespoň jedné letní školy
16	LETSKOLA_PO CET	1	počet účastí na letní škole
17	BUD_PREDSTAVA	A	student měl během studia představu o svém budoucím zaměstnání
18	ZAM_ZAMERENI	programování	pracovní zaměření studenta po absolvování
19	ZAM_SFERA	komerční	sféra zaměření studenta po absolvování
20	ZAM_FORMA	zaměstnanec	typ zaměstnání studenta po absolvování
21	ZAM_MESTO	Ostrava	město, ve kterém je absolvent zaměstnán
22	ZAM_PSC	70200	PSČ města, ve kterém je absolvent zaměstnán
23	ZAM_STAT	Česká republika	stát, ve kterém je absolvent zaměstnán
25	ZAM_NAZEV	Lufthansa Systems	společnost, ve které je absolvent zaměstnán
26	POHLAVI	M	pohlaví absolventa
27	OBEC	Hrádek	obec trvalého pobytu absolventa
28	PSC	73997	PSČ trvalého pobytu absolventa
29	STAT	Česká republika	stát trvalého pobytu absolventa
30	STUDIUM_ABS_BC_ROK	2010	rok absolvování bakalářského studia
31	STUDIUM_ABS_BC_UNI	UPOL	univerzita absolvování bakalářského studia
32	STUDIUM_ABS_MGR_ROK	2013	rok absolvování magisterského studia
33	STUDIUM_ABS_MGR_UNI	VŠB	univerzita absolvování magisterského studia
34	STUDIUM_ABS_PHD_ROK	2017	rok absolvování doktorského studia
35	STUDIUM_ABS_PHD_UNI	MUNI	univerzita absolvování doktorského studia
36	JMENO	Lucie	křestní jméno absolventa
37	PRIJMENI	NOVÁKOVÁ	příjmení absolventa
38	PRIJMENI_ROD	NOVOTNÁ	rodné příjmení absolventa

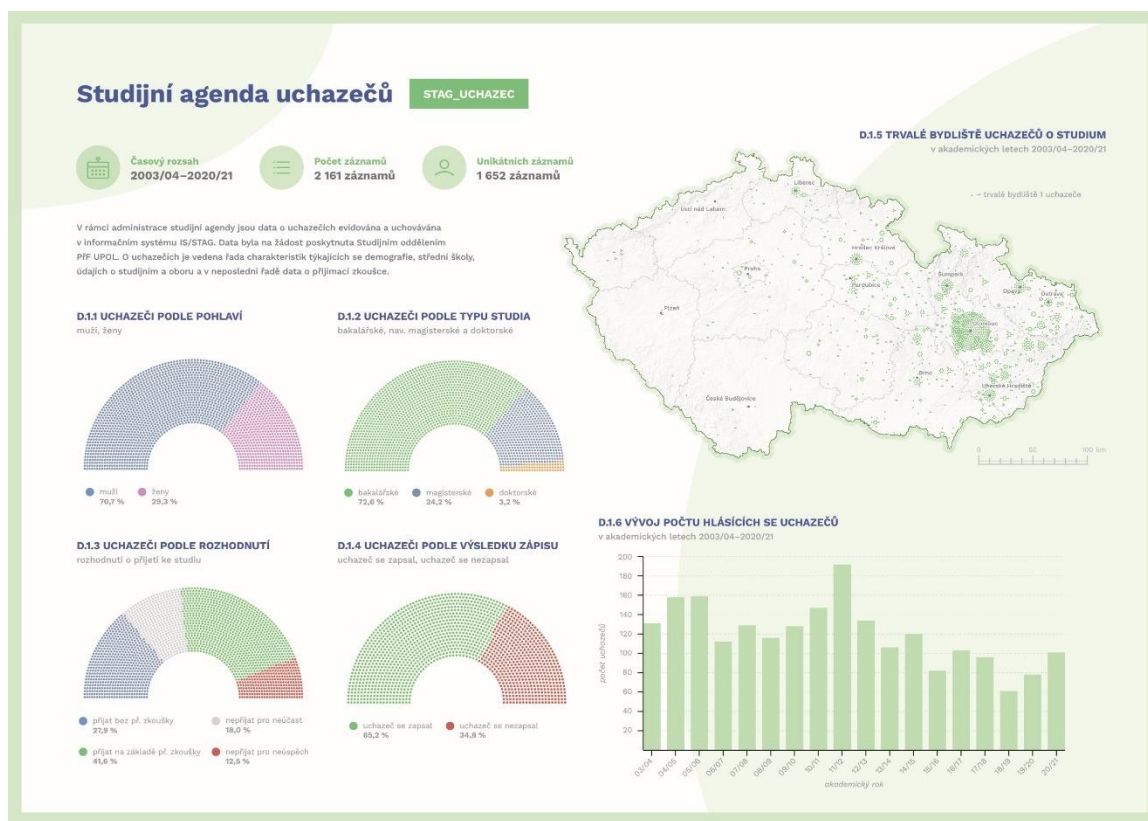
Příloha 1.6 – Přehled atributů datové sady KGI_STAZ

ID	Atribut	Příklad	Vysvětlení
1	ROK	<i>2004/2005</i>	akademický rok účasti na mobilitě
2	SEMESTR	<i>LS</i>	semestr účasti na mobilitě
3	PRIJMENI	<i>NOVÁK</i>	příjmení účastníka
4	KRESTNI	<i>Tomáš</i>	křestní jméno účastníka
5	JMENO	<i>NOVÁK Tomáš</i>	celé jméno účastníka
6	STAT	<i>Polsko</i>	navštívený stát
7	MESTO	<i>Lublaň</i>	navštívené město
8	UNIVERZITA	<i>Maria Curie-Sklodowska University</i>	navštívená univerzita
9	ROLE	<i>student</i>	role účastníka na mobilitě
10	PROGRAM	<i>CEEPUS</i>	program mobility

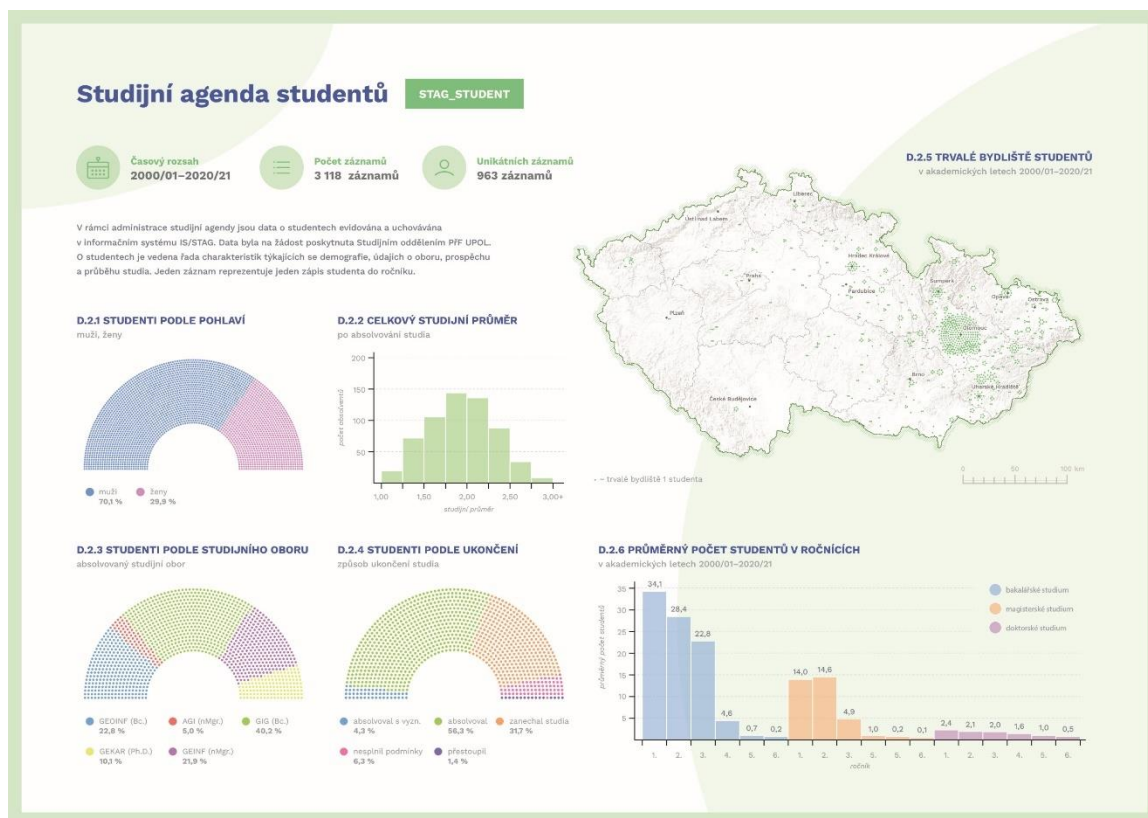
Příloha 1.7 – Přehled atributů datové sady KGI_UHAZEČ

ID	Atribut	Příklad	Vysvětlení
1	ROK	2016/2017	rok, do kterého se uchazeč hlásí
2	PRIJMENI	NOVÁK	příjmení uchazeče
3	JMENO	Tomáš	křestní jméno uchazeče
4	P_UCITELZ	ano	povědomí od učitele zeměpisu
5	P_UCITELI	ano	povědomí od učitele informatiky
6	P_PREDNASKA	ano	povědomí z přednášky na střední škole
7	P_GISDAY	ano	povědomí ze Dne GIS
8	P_DOD	ano	povědomí ze dne otevřených dveří
9	P_KAMARAD	ano	povědomí od kamaráda
10	P_RODIC	ano	povědomí od rodiče
11	P_GAUDE	ano	povědomí z veletrhu Gaudeamus
12	P_LETAK	ano	povědomí z tištěných materiálů
13	P_TV	ano	povědomí z TV
14	P_WWW	ano	povědomí z internetu
15	P_UCITNOV	ano	povědomí z učitelských novin
16	P_GISWEEK	ano	povědomí z akce GIS Weekend
17	NASTOUPIL	ano	výsledek nástupu uchazeče

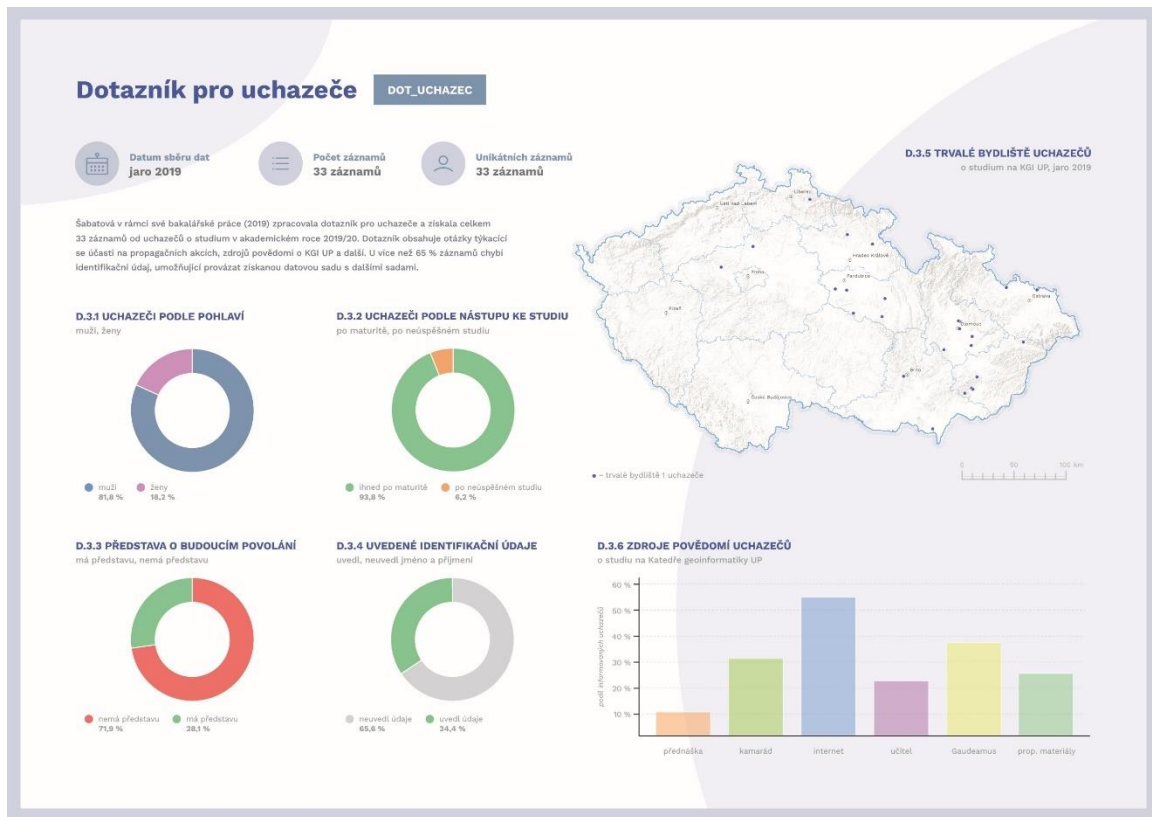
Příloha 3.1 – Náhled statistické infografiky **Studijní agenda uchazečů** (formát A3)



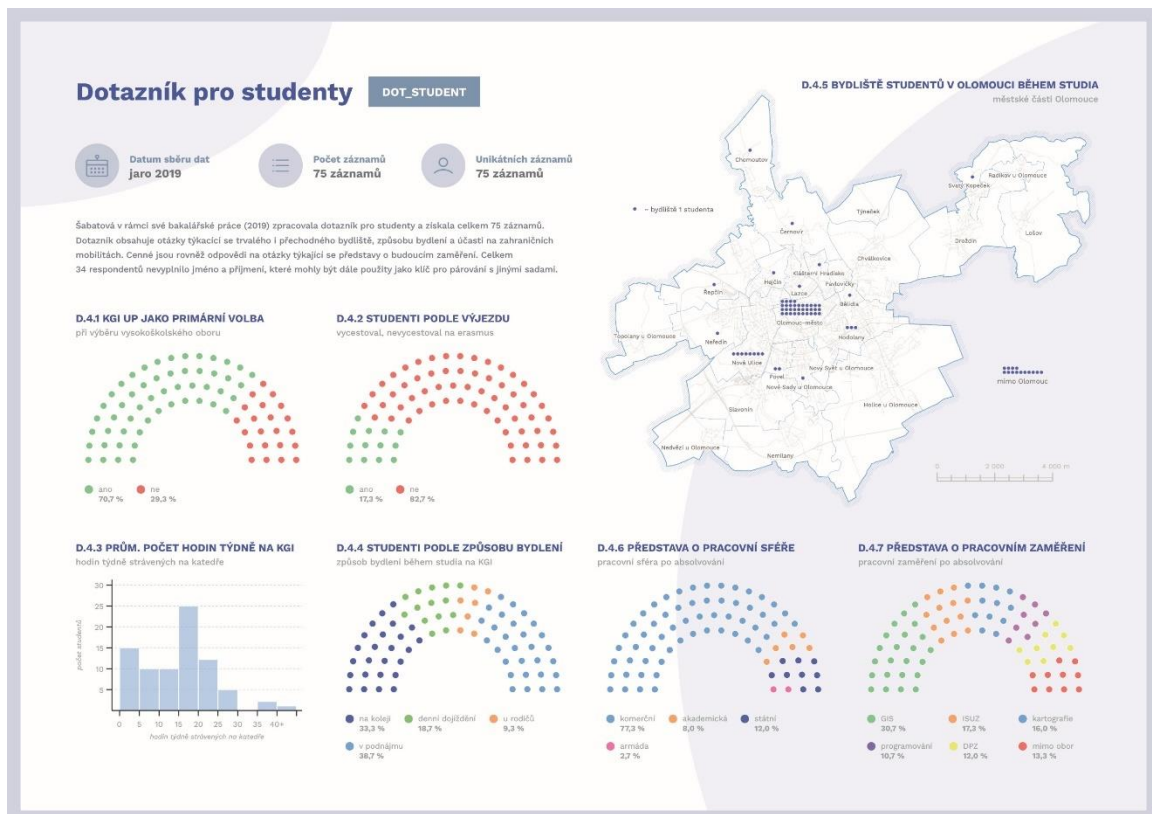
Příloha 3.2 – Náhled statistické infografiky **Studijní agenda studentů** (formát A3)



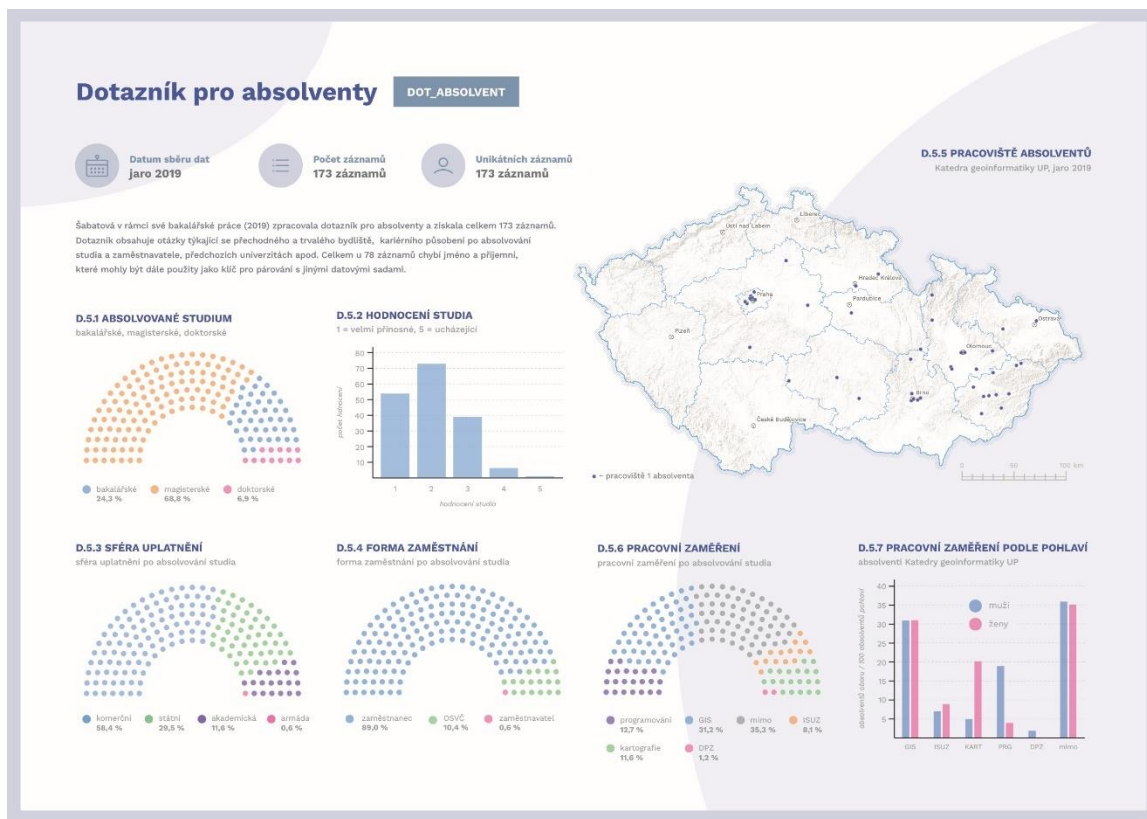
Příloha 3.3 – Náhled statistické infografiky **Dotazník pro uchazeče** (formát A3)



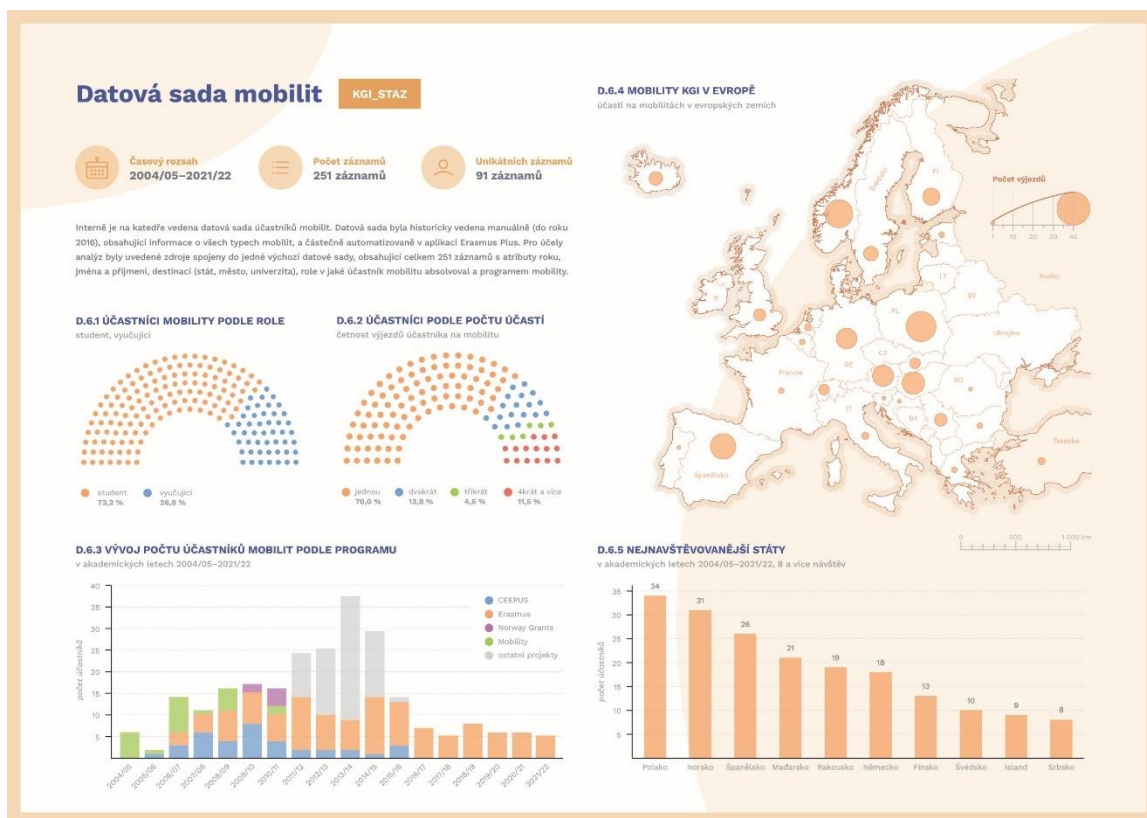
Příloha 3.4 – Náhled statistické infografiky **Dotazník pro studenty** (formát A3)



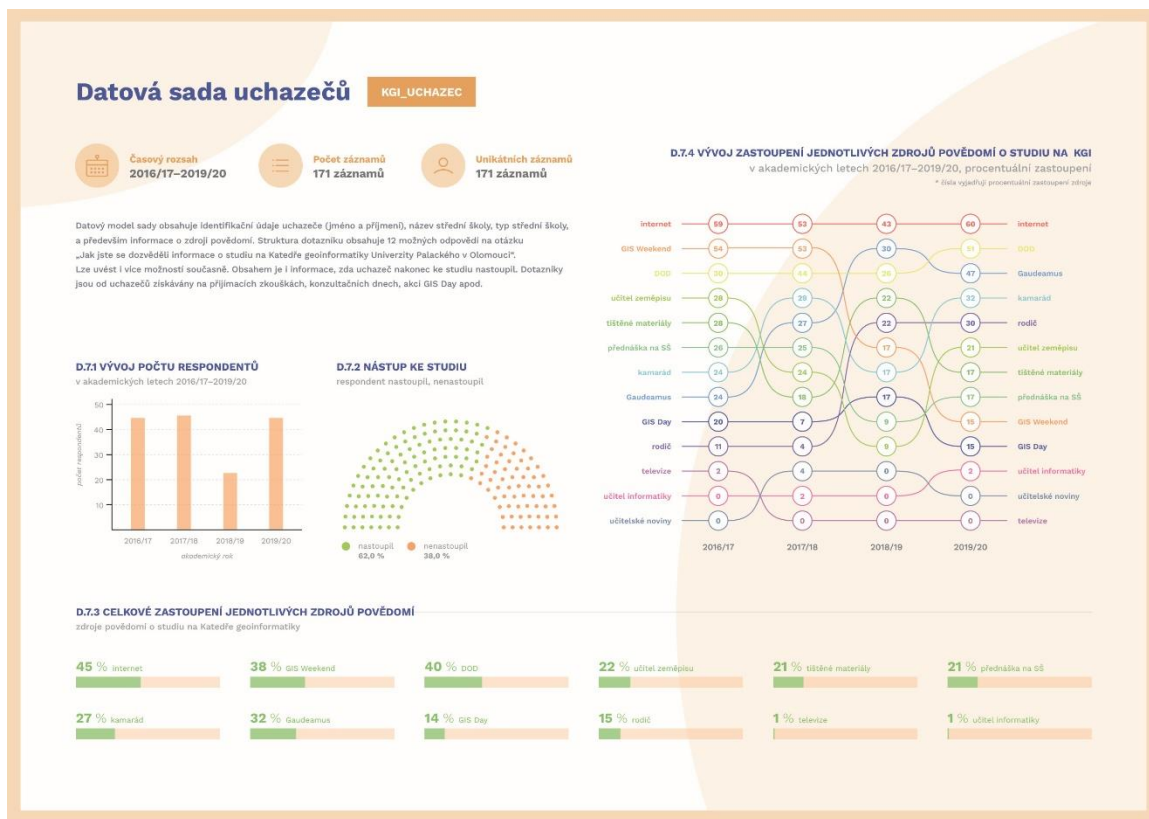
Příloha 3.5 – Náhled statistické infografiky **Dotazník pro absolventy** (formát A3)



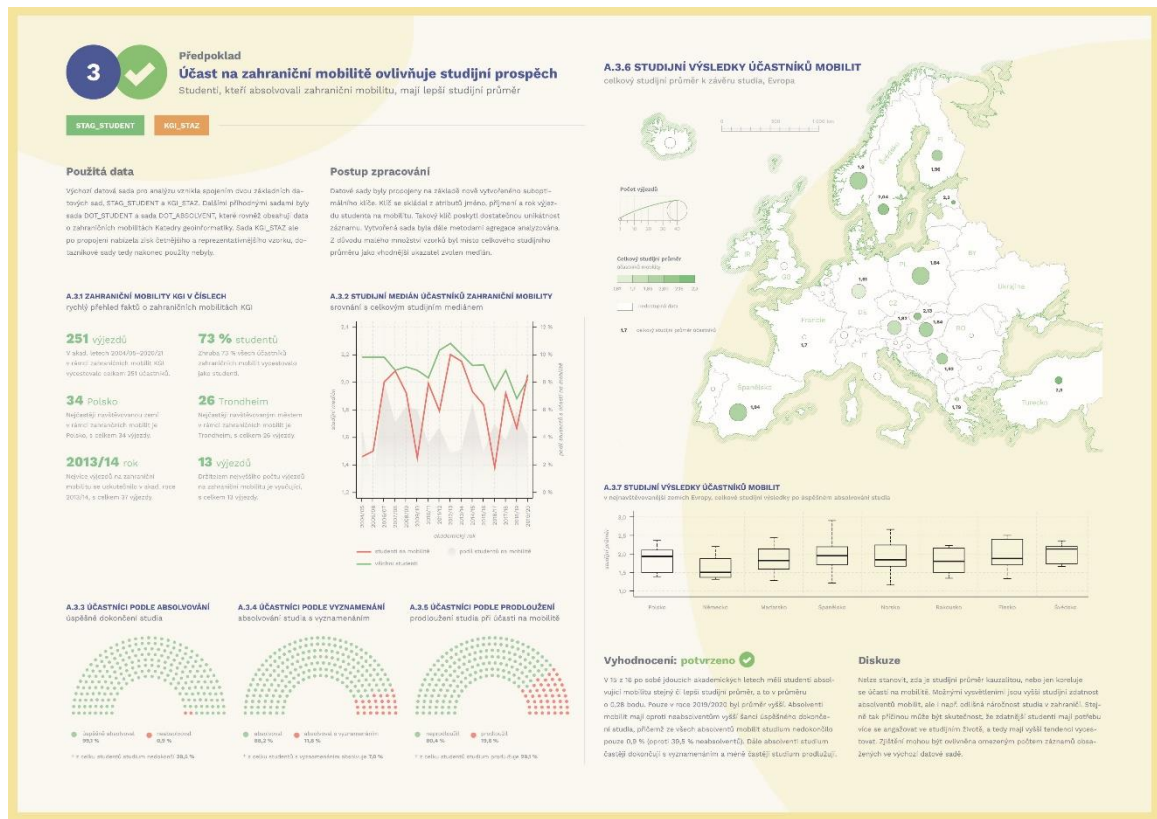
Příloha 3.6 – Náhled statistické infografiky **Datová sada mobilit** (formát A3)



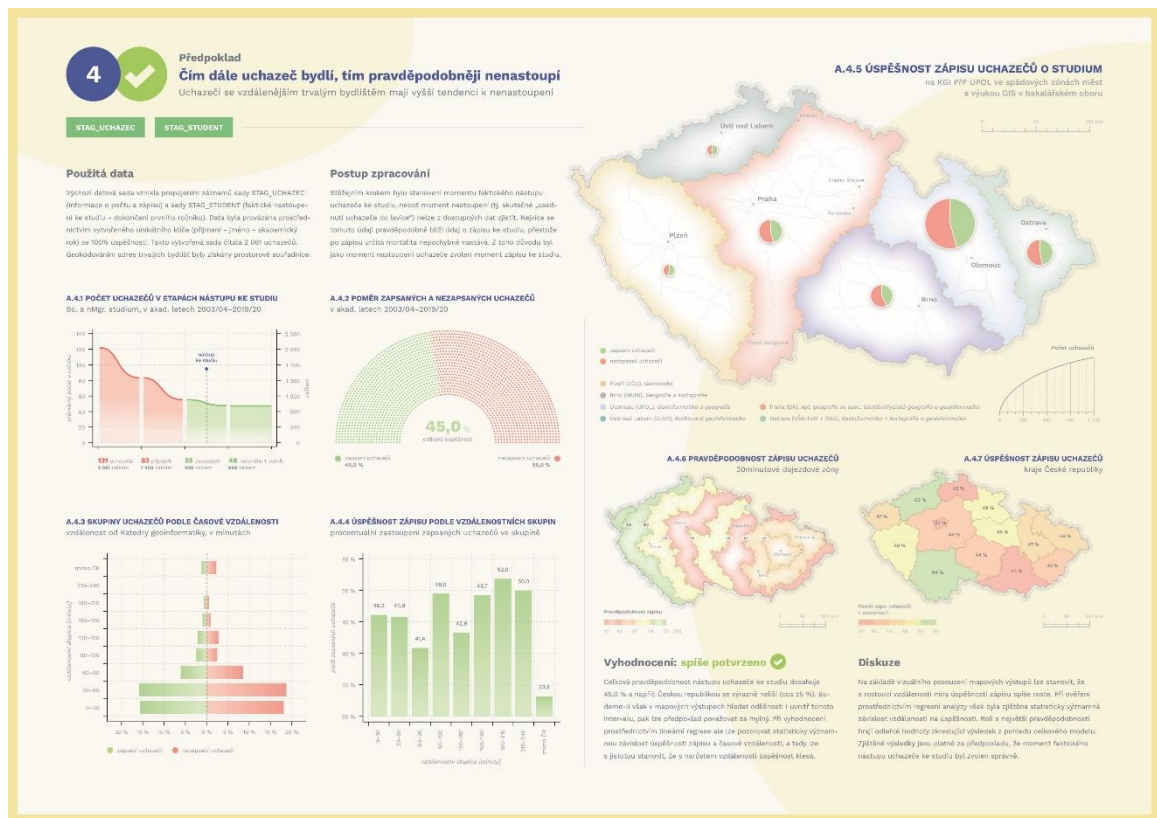
Příloha 3.7 – Náhled statistické infografiky **Datová sada uchazečů** (formát A3)



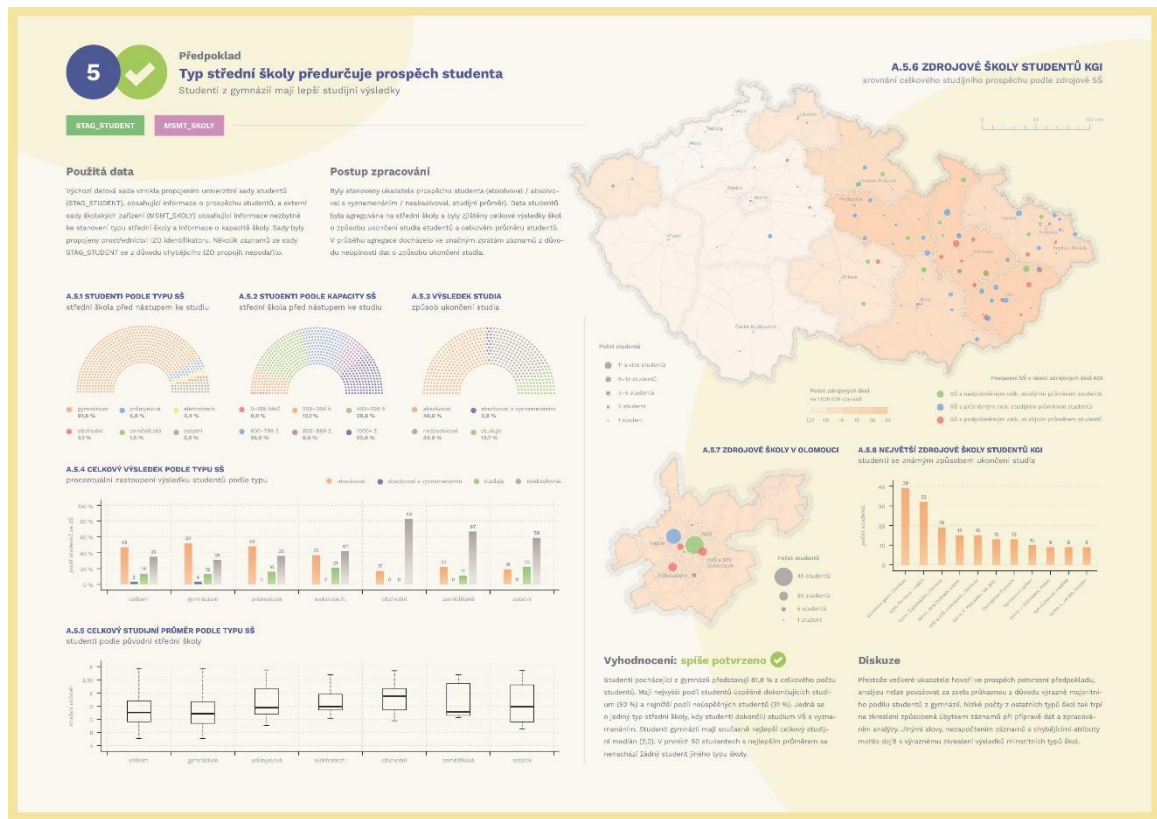
Příloha 4.3 – Náhled analytické infografiky **Předpoklad 3** (formát A2)



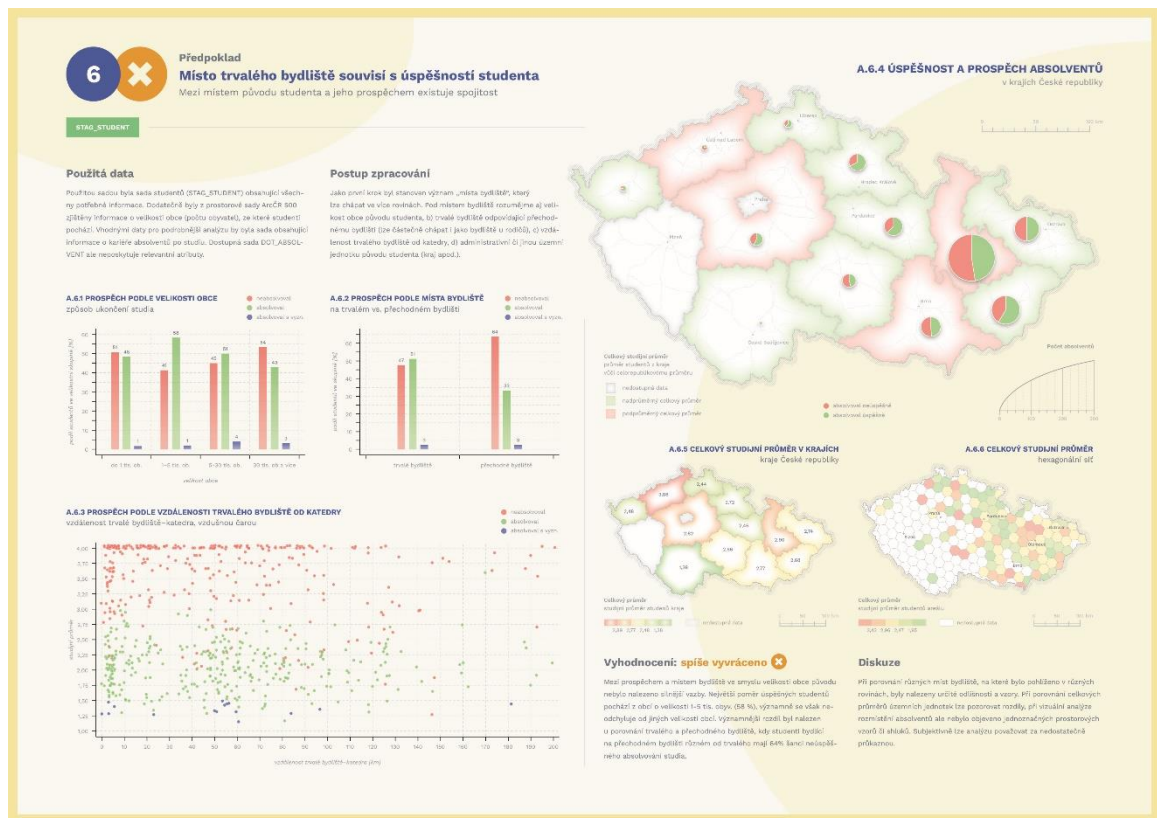
Příloha 4.4 – Náhled analytické infografiky **Předpoklad 4** (formát A2)



Příloha 4.5 – Náhled analytické infografiky **Předpoklad 5** (formát A2)



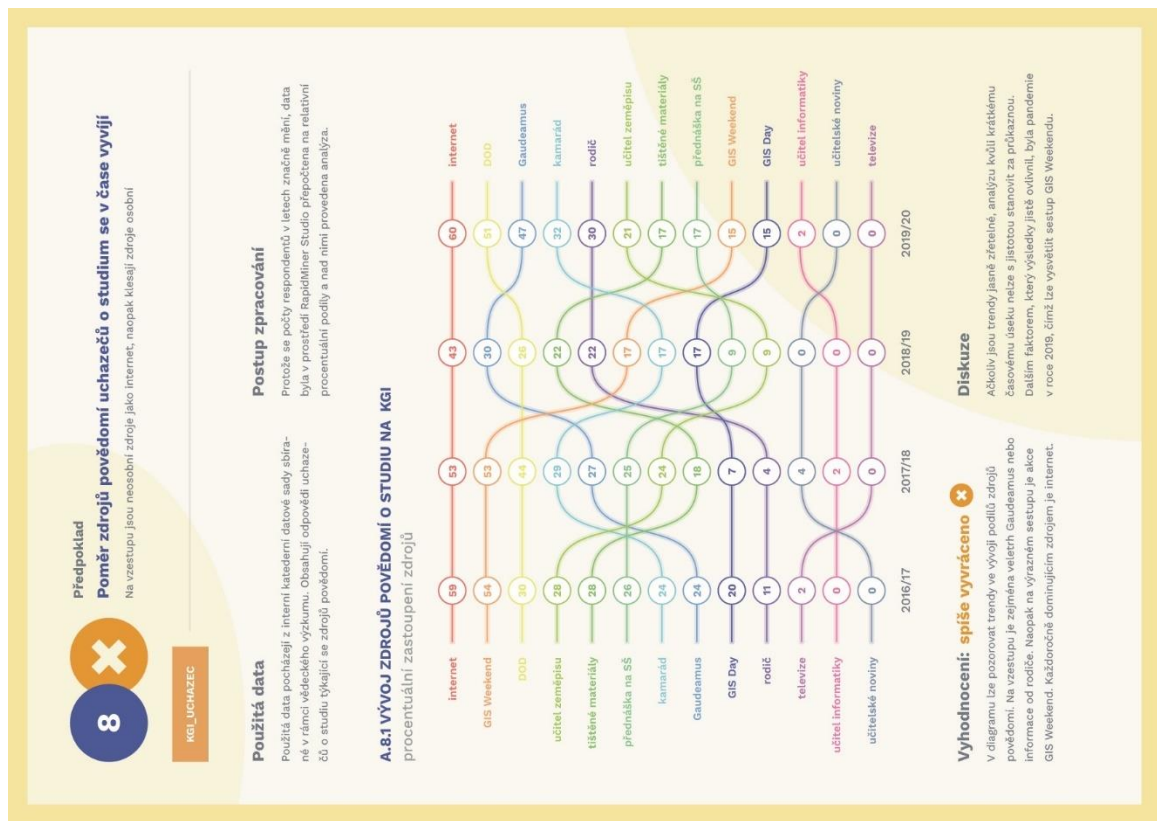
Příloha 4.6 – Náhled analytické infografiky **Předpoklad 6** (formát A2)



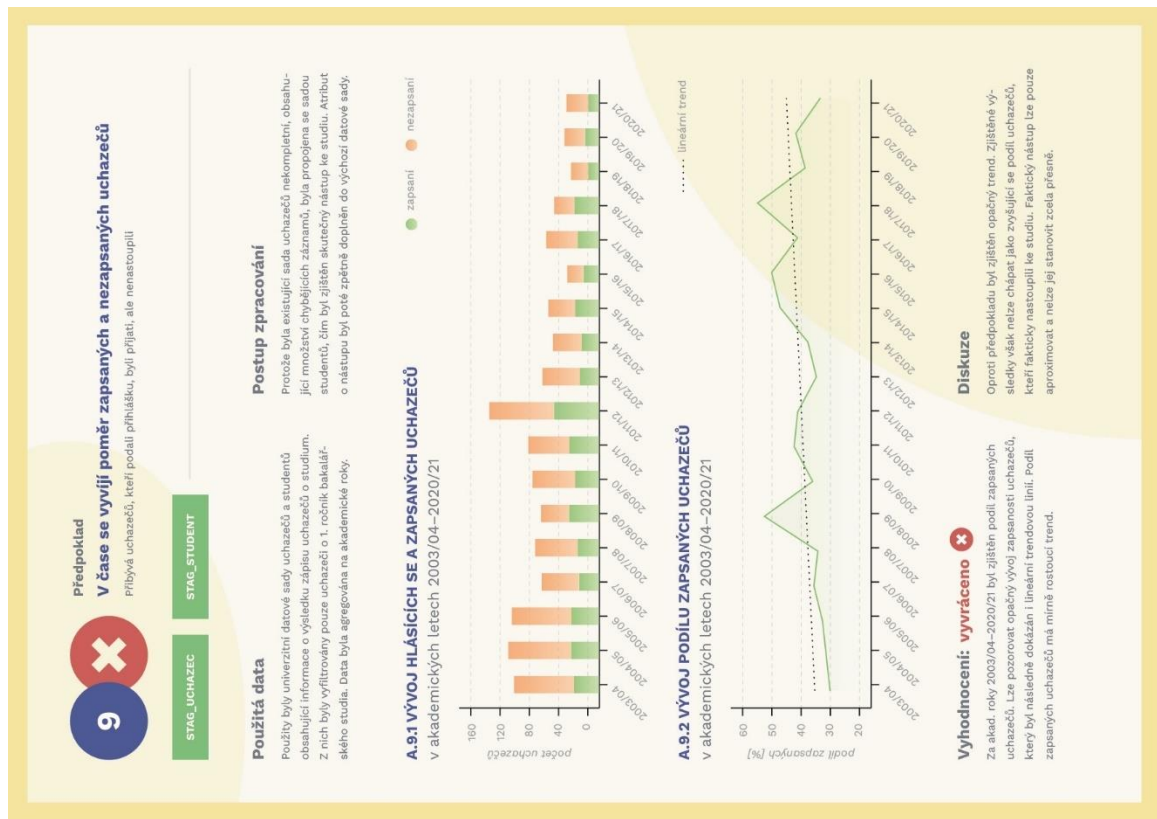
Příloha 4.7 – Náhled analytické infografiky Předpoklad 7 (formát A4)



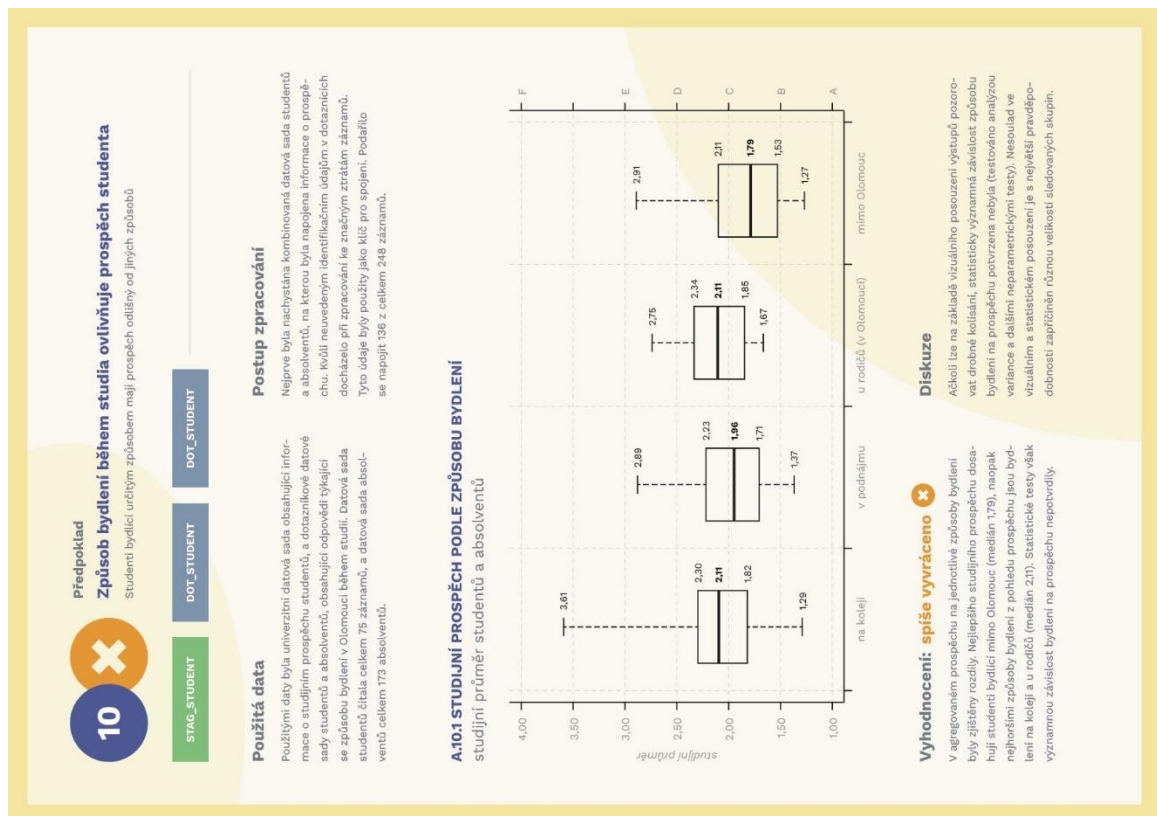
Příloha 4.8 – Náhled analytické infografiky Předpoklad 8 (formát A4)



Příloha 4.9 – Náhled analytické infografiky **Předpoklad 9** (formát A4)



Příloha 4.10 – Náhled analytické infografiky **Předpoklad 10** (formát A4)



Příloha 5 – Náhled webové prezentace (část 1)

Analýza studentů
Katedry geoinformatiky

O katedře Postup Data Statistika Analýza Závěr Ke stažení

Pokročilá analýza z dotazníkového šetření Katedry geoinformatiky

21 let historie 7 datových sad 12 analýz 1 web

Pojďme na to K analýze

O KATEDŘE

Katedra geoinformatiky

Katedra geoinformatiky Univerzity Palackého v Olomouci je univerzitní akademické pracoviště zabývající se geoinformačními systémy, kartografií, dálkovým průzkumem, modelováním prostorových jevů a příbuzných oborů. Katedra je členem v mnoha odborných skupinách a organizacích. Nabízí bakalářské, magisterské i doktorské studium.

Dálkový průzkum 3D tisk a modelování Drony a UAV Senzorové sítě Urbánní modelování

1 652 uchazečů
Mezi akademickými léty 2003/04–2020/21 se o studium ucházelo celkem 1 652 uchazečů.

963 studentů
Mezi akademickými léty 2000/01–2020/21 se ke studiu přihlásilo celkem 963 studentů.

656 absolventů
Mezi akademickými léty 2000/01–2020/21 úspěšně absolvovalo celkem 656 studentů.

POSTUP

Postup zpracování

Pokročilá analýza byla rozpracována do sedmi chronologicky na sebe navazujících fází. Během zpracování byl použit vodopádový přístup bez nutnosti retrospektivních kroků.

- Kompilace dat**
Nejprve byla shromážděna dostupná data obsahující informace o uchazečích, studentech a absolventech KGI.
- Transformace dat**
Ze sady byla vybrána data příhodná k analýze, očistěna o zbytečné atributy a transformována do potřebné podoby.
- Popisná statistika**
Seznámení se s jednotlivými datovými sadami proběhlo vypracováním jednoduchých statistických plakátů.
- Stanovení předpokladů**
S pomocí křížové matice byly stanoveny vhodné kombinace témat k posouzení. Následně byly stanoveny předpoklady.
- Pokročilá analýza**
Sešlacené předpoklady různými statistickými metodami prostorově

Příloha 5 – Náhled webové prezentace (část 2)

DATA

Použité datové sady

STAG_UCHAZEC

Studijní agenda uchazečů

👤 2 181 záznamů

📅 2003/04–2020/21

📄 Studijní odd.PFF UP

STAG_STUDENT

Studijní agenda studentů

👤 3 118 záznamů

📅 2000/01–2020/21

📄 Studijní odd.PFF UP

KGI_STAZ

Katederní sada mobilit

👤 251 záznamů

📅 2004/05–2021/22

📄 doc. Burian

KGI_UCHAZEC

Katederní sada uchazečů

👤 171 záznamů

📅 2016/17–2019/20

📄 doc. Dobešová

DGT_UCHAZEC

Dotazník pro uchazeče

👤 33 záznamů

📅 jaro 2019

📄 M. Šabatová

DGT_STUDENT

Dotazník pro studenty

👤 75 záznamů

📅 jaro 2019

📄 M. Šabatová

DGT_ABSOLVENT

Dotazník pro absolventy

👤 173 záznamů

📅 jaro 2019

📄 M. Šabatová

Identifikátor

Název datové sady

👤 počet záznamů

📅 datum sběru

📄 zdroj dat

zobrazit detail

STATISTIKA

Přehled dat

Primární soubor dohledaných dat souvisejících s katedrou obsahoval i další sady, často se však jednalo o souhrn či agregaci sad jiných. Proto byl zvolen užší výběr dat s největší informační hodnotou a nejvíce záznamy.

Celkem bylo vybráno 7 datových sad ze 3 různých zdrojů. Sady byly označeny šítky, které dále sloužily při tvorbě analýz k identifikaci použitých dat.

Přestože statistické zpracování bylo provedeno za účelem seznámení se s daty a vybudování přehledu ve výchozích materiálech, i to samotné přineslo mnohá zajímavá zjištění.

Studijní průměr

👤 **Celkový mediánový studijní průměr všech studentů je 2,26 (mezi známkou C a D).**

Dokončení s vyznamenáním

🎓 **S vyznamenáním absolvuje 4,3 % studentů, tzn. asi jeden z dvaceti.**

Poměr mužů a žen

👤 **Celkový poměr pohlaví studentů je 70,1 % mužů a 29,9 % žen.**

ANALÝZA

Pokročilá analýza dat

Na základě expertního názoru vybraných akademických pracovníků katedry a autorova uvážení bylo stanoveno celkem 12 předpokladů. Předpoklady byly ověřeny a vyhodnoceny.

1

PŘEDPOKLAD

Studenti se po absolvování vrací zpět do místa bydliště

Invalide bydliště před studiem a místo pracoviště po studiu by měly být podobné

→

PŘEDPOKLAD

Zdroj povědomí o studiu závisí na trvalém bydlišti uchazeče

Vzdělanější uchazeči dělájí povědomí o studiu jinými způsoby než uchazeči bydlící blíže

→

PŘEDPOKLAD

Účast na zahraniční mobilitě ovlivňuje studijní prospěch

Studenti, kteří absolvovali zahraniční mobilitu, mají lepší studijní průměr

→

PŘEDPOKLAD

Čím dále uchazeč bydlí, tím pravděpodobněji nenastoupí

Uchazeči se vzdálenějším trvalým bydlištěm mají větší tendenci k nenastoupení

→

PŘEDPOKLAD

PŘEDPOKLAD

PŘEDPOKLAD

PŘEDPOKLAD

Příloha 5 – Náhled webové prezentace (část 3)

Typ střední školy předurčuje prospěch studenta

Studenti z gymnázií mají lepší studijní výsledky

Místo trvalého bydliště souvisí s úspěšností studenta

Mezi místem původu studenta a jeho prospěchem existuje spojitost

Kariérní směr je předurčen pohlavím absolventa

Pohlaví ovlivňuje, jakou kariérní cestou, zaměřením a sférou se absolvent vydá

Poměr zdrojů povědomí uchazečů o studium se vlivy

Na vzestupu jsou neosobní zdroje jako internet, neopak klasických zdrojů osobní

9

V čase se vlivy poměr zapsaných a nezapsaných uchazečů

Přijevá uchazečů, kteří podali přihlášku, byli přijati, ale nenastoupili

10

Způsob bydlení během studia ovlivňuje prospěch studenta

Studenti bydlící určitým způsobem mají prospěch odlišný od jiných způsobů

11

Celková úspěšnost studentů se v čase vlivy

Vývoj mediánového průměru studentů má jasně zřetelný trend vývoje

12

Způsob bydlení ovlivňuje, zda student absolvoval mobilitu

Studenti, kteří nebyli v rodě, častěji vyjádřili na zahraniční mobilitu

Více informací o postupu a tvorbě analýz naleznete v textu práce.

[Soubory ke stažení](#)

PŘEHLED

Přehled předpokladů

<p>1 Studenti se po absolvování vrací zpět do místa bydliště</p> <p><small>Trvalé bydliště před studiem a místo pracoviště po studiu by měly být podobné</small></p> <p>2 Zdroj povědomí o studiu závisí na trvalém bydlišti uchazeče</p> <p><small>Vzdálenější uchazeči získali povědomí o studiu jinými způsoby než uchazeči bydlící blíže</small></p> <p>3 Účast na zahraniční mobilitě ovlivňuje studijní prospěch</p> <p><small>Studenti, kteří absolvovali zahraniční mobilitu, mají lepší studijní průměr</small></p> <p>4 Čím dále uchazeč bydlí, tím pravděpodobněji nenastoupí</p> <p><small>Uchazeči se vzdálenějším trvalým bydlištěm mají vyšší tendenci k nenastoupení</small></p> <p>5 Typ střední školy předurčuje prospěch studenta</p> <p><small>Studenti z gymnázií mají lepší studijní výsledky</small></p> <p>6 Místo trvalého bydliště souvisí s úspěšností studenta</p> <p><small>Mezi místem původu studenta a jeho prospěchem existuje spojitost</small></p>	<p>7 Kariérní směr je předurčen pohlavím absolventa</p> <p><small>Pohlaví ovlivňuje, jakou kariérní cestou, zaměřením a sférou se absolvent vydá</small></p> <p>8 Poměr zdrojů povědomí uchazečů o studium se vlivy</p> <p><small>Na vzestupu jsou neosobní zdroje jako internet, neopak klasických zdrojů osobní</small></p> <p>9 V čase se vlivy poměr zapsaných a nezapsaných uchazečů</p> <p><small>Přijevá uchazečů, kteří podali přihlášku, byli přijati, ale nenastoupili</small></p> <p>10 Způsob bydlení během studia ovlivňuje prospěch studenta</p> <p><small>Studenti bydlící určitým způsobem mají prospěch odlišný od jiných způsobů</small></p> <p>11 Celková úspěšnost studentů se v čase vlivy</p> <p><small>Vývoj mediánového průměru studentů má jasně zřetelný trend vývoje</small></p> <p>12 Způsob bydlení ovlivňuje, zda student absolvoval mobilitu</p> <p><small>Studenti, kteří nebyli v rodě, častěji vyjádřili na zahraniční mobilitu</small></p>
---	---

ZÁVĚR

Doporučení pro další sběr dat

Při kompletaci, přípravě a zpracování datových sad a analýz bylo zjištěno několik skutečností, které práci komplikovaly, zpomalovaly, nebo zapříčinily sníženou průkaznost výsledků analýz.

Existence těchto nedostatků je pochopitelná, neboť primárním účelem použitých datových souborů není a nebylo je takto analyzovat. S cílem usnadnění práce při dalších případných budoucích analýzách podobného charakteru byly posouzeny příčiny nedostatků a stanoveny doporučení pro další sběr dat a práci s nimi.

Doporučení není nutné brát dogmaticky. Vycházejí z osobní zkušenosti autora nabyté při zpracování analýz. Jejich cílem je urychlit práci s daty a v co největší možné míře zvýšit průkaznost analýz.

— Zamezení redundance atributů

Mezi datovými sadami byl zjištěn značný překryv v obsažených atributech. Atributy zřetelně vykytující se v jedné sadě lze snadno dohledat v sadě jiné, a pomocí různých klíčů propojit. Je proto vhodné dotazníky koncipovat tak, aby obsahovaly pouze otázky, na které nelze jiným způsobem zjistit odpovědi.

+ Kontinuální sběr dat

+ Identifikační údaje

+ Vnitřní členění dotazníků

Shrnutí...

Výsledkem práce je několik typů výstupů. Jejich přehled je následující.

7

datasetů

Byl analyzován a statisticky vizualizován datový soubor obsahující celkem 7 datových sad.

12

analýz

Bylo analyzováno 12 předpokladů. Výsledkem každé analýzy je přehledný plakát s jejich výsledky.

1


web

Hlavním výstupem práce je prezentace dosažených výsledků formou webové vizualizace.


Příloha 5 – Náhled webové prezentace (část 4)


KE STAŽENÍ

Soubory ke stažení





Text práce
PDF | 2,2 MB

Stáhnout 





Plakát
JPEG | 1,3 MB

Stáhnout 





Analýzy
ZIP | 12,3 MB

Stáhnout 



Přehled hodnocení
JPEG | 2,6 MB

Stáhnout 

 **katedra geoinformatiky**

AUTOR PRÁCE
Filip Fryčák
filip.frycak@gmail.com
Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

VEDOUcí PRÁCE
RNDr. Alena Vondráková, Ph. D., LL.M.
alena.vondrakova@upol.cz

KONZULTANT
Mgr. Jakub Konižek
jakub.konizek@upol.cz

© 2022 Filip Fryčák, Univerzita Palackého v Olomouci

Zpracováno jako příloha diplomové práce Pokročilá analýza a vizualizace z dotazníkového šetření Katedry geoinformatiky

Příloha 6 – Architektura projektu webové prezentace

```
root
|- assets/                                # zdrojové soubory: styly a funkcionalita
|   |- img/                               # obrázkové soubory
|   |   |- favicon/                      # favikony pro různé typy prohlížečů
|   |   |- icon/                         # SVG ikony
|   |   |- shapes/                      # SVG dekorační elementy
|   |   ...                              # další obrázkové soubory
|   |- js/                               # JavaScript soubory
|   |   |- vendors/                     # JavaScript knihovny
|   |   |- main.js                      # kořenový JavaScript soubory
|   |- scss/                             # SCSS soubory (blíže v příloze 7)
|   |   |- abstracts/                   # abstraktní styly nejvyšší úrovně
|   |   |- base/                        # základní styly
|   |   |- components/                 # komponenty
|   |   ...                              # další SCSS adresáře a soubory
|- data/                                  # zdrojové soubory: obsahová data
|   |- analyzy/                          # infografiky analýz (formát PNG a WEBP)
|   |- data/                             # infografiky dat (formát PNG a WEBP)
|   ...                                  # textová data (formátu JSON)
|- html/                                  # zdrojové soubory: strukturní NJK šablony
|   |- pages/                            # stránky (pouze index.njk)
|   |- templates/                       # strukturní šablony jednotlivých sekcí
|- node_modules/                         # vývojové JavaScript balíky
|- public/                                # distribuční kód
|   |- css/                              # kaskádové styly
|   |- data/                             # infografiky (formát PNG a WEBP)
|   |- fonts/                            # fonty
|   |- images                            # obrázkové soubory knihoven
|   |- img/                              # obrázkové soubory
|   |- js/                               # JavaScript kód
|   |- index.html                       # kořenový HTML soubor
|- .gitignore                            # konfigurační soubor systému Git
|- gulpfile.js                           # konfigurační soubor spouštěče úloh Gulp
|- package-lock.json                     # kořenový soubor vývojových balíků lock
|- package.json                          # kořenový soubor vývojových balíků
|- readme.md                             # dokumentace projektu
```

Příloha 7 – Architektura kaskádových stylů webové prezentace

```
scss/
|- abstracts/                                # abstraktní styly nejvyšší úrovně
|   |- _animations.scss                     # animace
|   |- _mixins.scss                         # funkcionální bloky
|   |- _reset.scss                          # anulování výchozích stylů
|   |- _shame.scss                          # nezařazené styly
|   |- _utilities.scss                      # jed noučelové styly
|   |- _variables.scss                      # SCSS proměnné
|- base/                                     # základní styly
|   |- _base.scss                           # základní styly
|   |- _button.scss                         # tlačítka
|   |- _card.scss                           # karty
|   |- _image.scss                          # obrázky
|   |- _tag.scss                            # tagy
|   |- _typography.scss                    # typografie a písmo
|- components/                              # komponenty
|   |- _accordion.scss                     # rolovací akordeony
|   |- _analysis-overview.scss             # přehled předpokladů
|   |- _analysis.scss                      # analytické karty
|   |- _data.scss                          # datové karty
|   |- _file.scss                           # soubory ke stažení
|   |- _result.scss                        # shrnutí
|   |- _slider-fact.scss                   # slider: stručná fakta
|   |- _slider-icon.scss                   # slider: katedra geoinformatiky
|   |- _slider-stat.scss                   # slider: statistiky
|   |- _slider-workflow.scss               # slider: postup zpracování
|   |- _socials.scss                       # odkazy na sociální sítě
|   |- _to-top.scss                        # tlačítko zpět nahoru
|- layout/                                  # rozvržení stránky
|   |- _footer.scss                        # patička
|   |- _header.scss                        # hlavička
|   |- _navbar.scss                        # navigace
|   |- _section.scss                       # sekce
|- vendors/                                 # přepis stylů třetích stran
|   |- _bootstrap.scss                     # přepis: bootstrap
|   |- _lightgallery.scss                  # přepis: lightgallery
|   |- _swiper.scss                        # přepis: swiper
|- main.scss                                # kořenový SCSS soubor
|- vendors.scss                             # styly třetích stran
```

Příloha 8 – Přehled vývojových JavaScript balíčků

Id	Balík	Verze	Popis (anglicky, zdroj: www.npmjs.com)
1	autoprefixer	10.4.4	PostCSS plugin to parse CSS and add vendor prefixes to CSS
2	bootstrap	5.1.3	Front-end framework for faster and easier web development
3	browser-sync	2.27.7	Time-saving synchronised browser testing
4	cssnano	5.1.3	A modular minifier, build on top of the PostCSS ecosystem
5	gulp	4.0.2	The streaming build system
6	gulp-autoprefixer	8.0.0	Prefix CSS (for Gulp)
7	gulp-concat	2.6.1	Concatenates files (for Gulp)
10	gulp-data	1.3.1	Proposes a common API for attaching data to the file object
11	gulp-nunjucks-render	2.2.3	Render Nunjucks templates with data (for Gulp)
12	gulp-postcss	9.0.1	PostCSS Gulp plugin (for Gulp)
13	gulp-rename	2.0.0	Rename files (for Gulp)
14	gulp-sass	5.1.0	Gulp plugin for SASS (for Gulp)
15	gulp-sourcemaps	3.0.0	Sourcemaps support for Gulp.js (for Gulp)
16	gulp-uglify	3.0.2	Minify files for Uglify.js (for Gulp)
17	jquery	3.6.0	JavaScript library for DOM operations
18	lg-zoom	1.3.0	Zoom module for lightgallery
19	lightgallery	2.4.0	Feature-rich, modular JavaScript gallery plugin
20	postcss	8.4.8	Tool for transforming styles with JS plugins
21	sass	1.49.9	A pure JavaScript implementation of SASS
22	swiper	8.0.7	Most modern mobile touch slider and framework