UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DETERMINANTY ÚNIKU MOZGOV DO ZAHRANIČIA

Diplomová práca

2024 **Bc. Samuel Hronský**

UNIVERZITA KOMENSKÉHO V BRATISLAVE

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

DETERMINANTY ÚNIKU MOZGOV DO ZAHRANIČIA

Diplomová práca

Študijný program:	Ekonomicko-finančná matematika a modelovanie
Študijný odbor:	1114 Aplikovaná matematika
Školiace pracovisko:	FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci:	doc. Mgr. Richard Kollár, PhD.

Bratislava 2024 **Bc. Samuel Hronský**



94080320

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Samuel Hronský
Študijný program: ekonomicko-finančná matematika a modelovanie (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Determinanty úniku mozgov do zahraničia
Determinants of international brain drain

Anotácia: Odchod študentov do zahraničia je celosvetový fenomén. O tom, ktoré faktory podmieňujú najviac cieľovú krajinu, sú však väčšinou len dohady. Cieľom práce je určiť z dostupných dát z databázy UNESCO dominantné faktory, ktoré určujú cieľovú krajinu študentov pri ich medzinárodnej vzdelávacej mobilite. Porovná sa aj to, či pre krajiny strednej Európy sú tieto faktory podobné, prípadne v nich existujú národné špecifiká.

Vedúci: doc. Mgr. Richard Kollár, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Marek Fila, DrSc.

Dátum zadania: 05.01.2023

Dátum schválenia: 11.01.2023

prof. RNDr. Daniel Ševčovič, DrSc.
garant študijného programu

.....
študent

.....
vedúci práce

Abstrakt

V tejto práci sa zaoberáme vplyvom vonkajších faktorov na migráciu študentov do zahraničia. Sústredíme sa pritom nielen na perspektívu slovenských študentov, ale aj na širší kontext vybraných krajín.

V úvodných častiach sa venujeme výberu krajín, pre ktoré budeme migráciu študentov analyzovať. Posúdime dotazníky vyplnené stredoškolskými študentmi a následne, s využitím týchto dotazníkov, konzultácií s odborníkmi a vlastnej intuície, vyberieme faktory, ktoré by mohli ovplyvniť rozhodnutie študentov migrovať do zahraničia. Každý z týchto faktorov krátko vysvetlíme a u zložitejších uvedieme postupy a metodológiu ich vytvorenia. Príkladom zložitého faktora je „Vážená vzdialenosť krajín“, pri ktorom predstavíme definíciu, odvodíme vzorec a opíšeme postup tvorby dát. Nie pre každú krajinu sa nám podarilo zozbierať všetky skúmané determinanty, čo nás vedie k vytvoreniu dvoch dátových súborov pre ďalšie analyzovanie.

V záverečnej časti preskúmame vplyv jednotlivých faktorov pomocou lineárnej regresie a rozhodovacích stromov. Postupy analyzovania rozdelíme na lokálny a globálny model. V rámci lokálneho modelu budeme analyzovať lineárnu regresiu pre jednotlivé krajiny od samého začiatku. V globálnom modeli najprv identifikujeme signifikantné faktory a potom vytvoríme globálny regresný model, ktorý bude platný pre všetky krajiny vzorky. Nakoniec, pri rozhodovacích stromoch, analyzujeme dôvody migrácie študentov podľa krajín, aby sme pochopili ich motiváciu migrovať do zahraničia. Naším cieľom je poskytnúť schopnosť interpretovať kľúčové aspekty migrácie pre študentov z konkrétnych krajín.

Abstract

In this work, we examine the impact of external factors on students' migration abroad. We focus not only on the perspective of Slovak students but also on the broader context of selected countries.

In the introductory sections, we address the selection of countries for which we will analyze student migration. We will assess questionnaires filled out by high school students and then, using these questionnaires, consultations with experts, and our own intuition, we will select factors that could influence the decision of students to migrate abroad. We will briefly explain each of these factors and, for more complex ones, present the procedures and methodology for their creation. An example of a complex factor is the 'Weighted distance of countries', for which we will introduce the definition, derive the formula, and describe the data creation process. We were not able to collect all the studied determinants for every country, leading us to create two datasets for further analysis.

In the final part, we will explore the impact of individual factors using linear regression and decision trees. We will divide the analysis procedures into a local and a global model. Within the local model, we will analyze linear regression for individual countries from the beginning. In the global model, we first identify significant factors and then create a global regression model that will be valid for all countries in the sample. Finally, with decision trees, we analyze the reasons for students' migration by country to understand their motivation to migrate abroad. Our goal is to provide the ability to interpret key aspects of migration for students from specific countries.

Poďakovanie

Rád by som vyjadril svoju hlbokú vďačnosť a uznanie všetkým, ktorí mi pomohli pri tvorbe tejto diplomovej práce. Osobitné poďakovanie patrí môjmu školiteľovi, pánovi Richardovi Kollárovi, za jeho neoceniteľnú podporu, cenné rady a odborné vedenie počas celého procesu prípravy práce. Jeho odborné znalosti a osobný prístup boli pre mňa neustálou inšpiráciou a motiváciou.

Taktiež by som chcel vyjadriť vďaku pánom Miloslavovi Bahnovi a Martinovi Šusterovi, ktorí ako odborní konzultanti významne prispeli k odbornosti a kvalite mojej práce. Ich odborné pripomienky a konštruktívna kritika boli pre mňa veľkým prínosom a pomohli mi lepšie pochopiť komplexné aspekty mojej témy a doplniť chýbajúce parametre v ich oblastiach pôsobenia.

Ďalej by som chcel vyjadriť svoju vďačnosť pánovi Mauriziovi Servovi za jeho láskavosť a ochotu sprístupniť dáta Levensteinových vzdialeností indoeurópskych jazykov, ktoré boli kľúčové pre vytvorenie parametra jazykových vzdialeností krajín.

Výraznú vďačnosť dlžím aj spoločnosti Finax, o.c.p. a.s., za poskytnutú príležitosť využívať inovatívne technológie, ako sú Open AI a Power BI, ktoré boli kľúčovými nástrojmi v mojej práci. Ich podpora bola neoceniteľná nielen pre praktickú časť mojej práce, ale aj pre rozšírenie mojich poznatkov a skúseností v tejto oblasti.

Nakoniec, ale rozhodne nie posledne, chcem vyjadriť svoju vďačnosť svojej rodine a blízkym za ich nekonečnú podporu, trpezlivosť a povzbudzovanie, ktoré mi boli oporou počas celého štúdia. Bez ich morálnej a emocionálnej podpory by som nemohol dosiahnuť to, čo mám dnes.

Obsah

1 Výber krajín **10**

1.1 Výber krajín - prvá skupina dát 11

1.2 Výber krajín - druhá skupina dát 12

1.3 Krajiny mimo EU 12

2 Dotazník študentov stredných škôl **13**

3 Opis parametrov **15**

3.1 Vzdelávanie 15

3.1.1 Rebríček vzdelávania 15

3.1.2 Vládne výdavky na vysokoškolské vzdelávanie ako % HDP a Vládne výdavky na vysokoškolské vzdelávanie,v PKS v miliónoch $ 16

3.1.3 Svetový rebríček univerzít 2023 17

3.1.4 Hrubý pomer ukončenia vysokoškolského vzdelávania 18

3.1.5 Výdavky na vývoj a výskum ako percento z HDP 19

3.2 Ekonomické parametre 19

3.2.1 Parametre spojené s HDP 19

3.2.2 Priemerný príjem PPP 20

3.2.3 Giniho index 2020 20

3.2.4 Zamestnanosť a nezamestnanosť 21

3.2.5 Inflácia 21

3.3 Sociálne parametre 23

3.3.1 Politická stabilita a absencia násilia/terorizmu odhad 23

3.3.2 Dĺžka života 24

3.3.3 Rebríček korupcie 24

3.3.4 Jazykové vzdialenosti krajín na základe Levensteinových vzdialeností . . . 24
3.3.5 Etická, jazyková a náboženská frakcionalizácia . . . 26
3.3.6 Index lepšieho života (Better Life Index) . . . 26
3.3.7 Minulé toky študentov . . . 26
3.4 Demografické parametre . . . 27
3.4.1 Susedné štáty . . . 27
3.4.2 Populácia . . . 27
3.5 Vzdialenosť dvoch krajín - Vážená geografická vzdialenosť a vážená časova vzdialenosť cesty autom . . . 28
3.5.1 Návrh výpočtu . . . 30
3.5.2 Implementácia a praktická aplikácia odvodeného vzorca . . . 31
3.5.3 Porovnanie vzdialeností hlavných miest . . . 32
3.6 Svetový rebríček konkurencieschopnosti v digitalizácii krajiny 2018 . . . 36

4 Lineárna regresia **37**

4.1 Modely . . . 37
4.2 Vytvorenie modelu po krajinách z prvej skupiny dát . . . 39
4.3 Vytvorenie globálneho modelu prvej skupiny dát . . . 44
4.3.1 Vysvetlenie parametra Minulé toky študentov . . . 53
4.3.2 Vytvorenie submodelu s jazykovými vzdialenosťami . . . 56
4.4 Vytvorenie globálneho modelu druhej skupiny dát . . . 57

5 Rozhodovacie stromy **58**

5.1 Rozhodovacie stromy - prvá skupina dát . . . 59
5.2 Rozhodovacie stromy - druhá skupina dát . . . 65

6 Použitie umelej inteligencie (UI) pri tvorbe práce **71**

7 Záver **73**

8 Prílohy **74**

Úvod

V našej diplomovej práci sa venujeme komplexnému fenoménu známemu ako „únik mozgov“, ktorý znamená migráciu študentov z jednej krajiny do druhej. Tento trend má významný vplyv nielen na jednotlivé krajiny, ale aj na globálnu ekonomiku a spoločnosť. Naším cieľom je preskúmať a identifikovať kľúčové determinanty ovplyvňujúce tento proces, so zameraním na faktory ako vzdelávacie príležitosti, ekonomické perspektívy, politickú stabilitu, sociálne faktory a demografické ukazovatele.

Na túto problematiku sa zameralo viacero sociológov a vedcov. Štúdia o amerických študentoch študujúcich v zahraničí [4], poskytuje hlbší pohľad na dynamiku a motiváciu, ktoré vedú študentov k rozhodnutiu emigrovať za vzdelaním alebo prácou. Tento výskum ukázal, že študenti s vyšším záujmom o štúdium v zahraničí majú tendenciu mať vyššiu úroveň motivácie, neofíliu, „migrantskú osobnosť“ a chuť pomáhať ostatným. Avšak, štúdia sa zameriava len na amerických študentov, čo sa môže líšiť od motivácie študentov napríklad nórskeho pôvodu študovať v zahraničí. My predpokladáme, že modely odlevu študentov v jednotlivých krajinách sa môžu líšiť, avšak nevylučujeme ani existenciu globálneho modelu, ktorý sa tiež budeme snažiť zostrojiť v ďalších častiach práce.

Ďalším príspevkom je práca [1]. Štúdia „Migrácia študentov v Európe: súťaž o ľudský kapitál“ sa zameriava na migračné toky európskych vysokoškolských študentov a odhaľuje, že sú ovplyvnené faktormi ako kultúrna podobnosť, jazykové schopnosti, obchodné toky a medzinárodná spolupráca v oblasti vzdelávania. Výsledky potvrdzujú prirodzené očakávania. Kultúrna podobnosť, najmä jazyk, zohráva kľúčovú úlohu pri ovplyvňovaní vzorcov migrácie študentov. Štúdia naznačuje, že faktory ako spoločný jazyk významne formujú tok poznatkov a študentov. Metódy práce obsahovali určité limitácie, ktoré sa budeme snažiť odstrániť. Baláž využíva len dvojfaktorovú analýzu vplyvov, ale v reálnom svete sa parametre navzájom prelínajú a ich vplyv často nie je lineárny. Našim cieľom je využiť moderné technológie, umelú inteligenciu a konzultácie s ľuďmi z praxe na hlbšie preskúmanie tejto problematiky a na vytvorenie viacerých parametrov, ktoré nie sú voľne dostupné na internete. V percepcii študentov často vystupujú nemerateľné alebo ťažko merateľné parametre.

V práci používame rôzne výskumné metódy, vrátane analýzy sekundárnych dát,

prieskumov a prípadových štúdií, analýza rozhodovacích stromov a lineárnu regresiu aby sme poskytli komplexný obraz o determinantoch úniku mozgov. Očakávame, že ako signifikantný sa ukáže vplyv, niektorých „nečakaných parametrov.“ Tento multidisciplinárny prístup nám umožňuje lepšie pochopenie toho, ako rôzne faktory interagujú a ovplyvňujú rozhodnutia jedincov, a zároveň ponúka užitočné návrhy pre tvorcov politík a vzdelávacie inštitúcie, ako ovplyvniť migráciu.

V práci sa zameriavame na skúmanie úniku mozgov z pohľadu viacerých krajín Európskej únie. Snažíme sa získať podrobný pohľad na proces rozhodovania sa náhodného študenta v každej z týchto krajín a opísať motivácie študentov krajiny Y pre štúdium v krajine Z. Naším cieľom je poskytnúť detailný pohľad na dôvody, prečo študenti opúšťajú svoju domovinu. Táto práca prispieva k lepšiemu porozumeniu tohto komplexného a multidimenzionálneho javu, ktorý je dôležitý pre rozvoj udržateľných a efektívnych stratégií v oblasti migrácie a vzdelávania.

Ako sme preukázali v našej práci, proces rozhodovania študentov sa líši v závislosti od ich štátnej príslušnosti. Naša práca tiež poukazuje na nedostatok údajov zbieraných vo svetových krajinách vrátane Európy. Odvodili sme tiež parametre, ktoré sú náročné na interpretáciu, ako napríklad vážené vzdialenosti, jazykové vzdialenosti a tabuľky odzrkadľujúce rôzne stupne kvality vzdelávania v krajinách. Na základe dátovej analýzy sa nám podarilo preukázať trendy migrácie v rámci Európy a vytvoriť globálny model migrácie študentov z európskych krajín. V poslednej časti práce predstavíme vytvorenie rozhodovacích stromov a poukážeme na ich rozdielnosť v závislosti od krajín pôvodu študentov.

1 Výber krajín

Pôvodne sme zamýšľali vykonať analýzu migračných tokov v rámci Európy s cieľom dosiahnuť čo najvyššiu flexibilitu využitia tejto práce. Už pri počiatočnom triedení dát týkajúcich sa odlivu študentov medzi krajinami sme narazili na niektoré štáty s chýbajúcimi údajmi v sledovanom období 2016–2019. V prvej iterácii sme preto vylúčili tieto krajiny: Gibraltár, Vatikán a Čierna Hora.

V priebehu práce sme identifikovali ďalšie chýbajúce údaje pre niektoré zo zvyš-

ných krajín. Postupne sme tak z našej sady európskych krajín vynechali Albánsko, Moldavsko, Lichtenštajnsko, Monako a San Marino.

Je známe, že európski študenti do značnej mieri migrujú aj mimo kontinent Európy. Rozhodli sme sa preto rozšíriť súbor DESTINATIONS, ktorý zahŕňa krajiny, kam európski študenti migrujú, o nasledujúce svetové veľmoci: Spojené štáty americké, Kanada, Austrália, Japonsko, Južná Kórea, Čína, Singapur, Nový Zéland, Spojené arabské emiráty, Brazília, Mexiko, Argentína, Južná Afrika, India a Turecko. Aj u niektorých mimoeurópskych krajín sme však narazili na chýbajúce údaje pre kľúčové parametre, čo nás nakoniec viedlo k vylúčeniu Číny, Mexika, Argentíny a Singapuru.

Napriek pokročilým technológiám dnešného sveta má mnoho krajín problémy so zberom údajov, a tiež sme narazili na problémy s relevanciou dát. Väčšina zaujímavých parametrov bola zbieraná len v rámci skupín krajín. Najviac takýchto údajov sme našli pre krajiny OECD, napríklad Better Life Index so škálou rôznych ťažko meratelných parametrov. Preto sme sa nakoniec rozhodli vytvoriť dve skupiny dát, ktoré budú základom pre modely v našej práci. Prvá skupina dát bude obsahovať všetky nevylúčené krajiny z predchádzajúcej pasáže spolu so zvyšnými mimoeurópskymi krajinami, pre ktoré sme našli dostatočné množstvo údajov. Druhá skupina dát bude špeciálne určená iba pre krajiny OECD a bude obohatená o viacero parametrov.

1.1 Výber krajín - prvá skupina dát

Počiatočné krajiny z ktorých nás zaujíma migracia študentov (ORIGINS): *Belgicko, Bielorusko, Bosna a Hercegovina, Bulharsko, Chorvátsko, Česko, Dánsko, Estónsko, Francúzsko, Fínsko, Grécko, Holandsko, Island, Írsko, Litva, Lotyšsko, Luxembursko, Malta, Maďarsko, Nemecko, Nórsko, Portugalsko, Poľsko, Rakúsko, Rumunsko, Rusko, Severné Macedónsko, Slovensko, Slovinsko, Spojené kráľovstvo, Srbsko, Španielsko, Švajčiarsko, Švédsko, Taliansko, Ukrajina.*

Krajiny do ktorých nás zaujíma prílev študentov (DESTINATIONS): *Austrália, Belgicko, Bielorusko, Bosna a Hercegovina, Brazília, Bulharsko, Chorvátsko, Česko, Dánsko, Estónsko, Francúzsko, Fínsko, Grécko, Holandsko, India, Island, Írsko, Japonsko, Južná Afrika, Južná Kórea, Kanada, Litva, Lotyšsko, Luxembursko, Malta, Maďarsko, Nemecko, Nový Zéland, Nórsko, Portugalsko, Poľsko, Rakúsko, Rumunsko,*

Rusko, Severné Macedónsko, Slovensko, Slovinsko, Spojené arabské emiráty, Spojené kráľovstvo, Spojené štáty americké, Srbsko, Španielsko, Švajčiarsko, Švédsko, Taliansko, Turecko, Ukrajina.

1.2 Výber krajín - druhá skupina dát

Počiatočné krajiny z ktorých nás zaujíma migracia študentov (ORIGINS): *Belgicko, Česko, Dánsko, Estónsko, Francúzsko, Fínsko, Grécko, Holandsko, Island, Írsko, Litva, Lotyšsko, Luxembursko, Maďarsko, Nemecko, Nórsko, Portugalsko, Poľsko, Rakúsko, Rusko, Slovensko, Slovinsko, Spojené kráľovstvo, Španielsko, Švajčiarsko, Švédsko, Taliansko.*

Krajiny do ktorých nás zaujíma prílev študentov (DESTINATIONS): *Austrália, Belgicko, Brazília, Česko, Dánsko, Estónsko, Francúzsko, Fínsko, Grécko, Holandsko, Island, Írsko, Japonsko, Južná Afrika, Južná Kórea, Kanada, Litva, Lotyšsko, Luxembursko, Maďarsko, Nemecko, Nový Zéland, Nórsko, Portugalsko, Poľsko, Rakúsko, Rusko, Slovensko, Slovinsko, Spojené kráľovstvo, Spojené štáty americké, Španielsko, Švajčiarsko, Švédsko, Taliansko, Turecko.*

1.3 Krajiny mimo EU

Spojené štáty - USA sú domovom niektorých najlepších univerzít na svete a ponúkajú širokú škálu akademických programov v rôznych disciplínach, čo z nich robí hlavnú destináciu pre medzinárodných študentov.

Kanada - Kanada je známa svojou vysokou kvalitou vzdelania a multikultúrnym prostredím, čo ju robí atraktívnou destináciou pre medzinárodných študentov.

Austrália - Austrália má silný vzdelávací systém a ponúka uvoľnený životný štýl a krásnu prírodnú scenériu, čo ju robí populárnou destináciou pre študentov, ktorí hľadajú jedinečnú skúsenosť so štúdiom v zahraničí.

Japonsko - Japonsko je lídrom v technológiách a inováciách, ako aj krajinou s jedinečnou kultúrou a históriou, čo ju robí atraktívnou destináciou pre študentov študujúcich obory ako inžinierstvo, podnikanie a japonský jazyk a kultúru.

Južná Kórea - Južná Kórea je rýchlo sa rozvíjajúcou ekonomikou so silným zameraním na inovácie a technológie, čo ju robí atraktívnou destináciou pre študentov študujúcich STEM predmety.

Nový Zéland - Nový Zéland je známy pre svoju prírodnú krásu a životný štýl v prírode, ako aj pre svoj vysokokvalitný vzdelávací systém, čo ho robí atraktívnou destináciou pre študentov, ktorí hľadajú jedinečnú skúsenosť so štúdiom v zahraničí.

Spojené arabské emiráty - UAE je rýchlo sa rozvíjajúcou krajinou s silnou ekonomikou a multikultúrnou spoločnosťou, čo z neho robí atraktívnu destináciu pre študentov, ktorí chcú študovať na Blízkom východe.

Brazília - Brazília je najväčšou krajinou Južnej Ameriky s živou kultúrou a ekonomikou, čo z nej robí atraktívnu destináciu pre študentov študujúcich oblasti, ako sú podnikanie, jazyk a kultúra a environmentálne štúdie.

Južná Afrika - Južná Afrika je krajinou s bohatou históriou a rôznorodou kultúrou, ako aj s rastúcou ekonomikou a vynikajúcimi univerzitami, čo z nej robí atraktívnu destináciu pre študentov študujúcich oblasti, ako sú spoločenské vedy, environmentálne štúdie a africké štúdie.

India - India je rýchlo sa rozvíjajúca krajina s bohatým kultúrnym dedičstvom a rastúcou ekonomikou, čo ju robí atraktívnym cieľom pre študentov študujúcich oblasti ako podnikanie, strojárstvo a indický jazyk a kultúru.

Turecko - Turecko je krajina s bohatou históriou a kultúrou, ako aj rýchlo sa rozvíjajúcou ekonomikou a vynikajúcimi univerzitami, čo ju robí atraktívnym cieľom pre študentov študujúcich oblasti ako história, jazyk a kultúra a politické vedy.

2 Dotazník študentov stredných škôl

V rámci našej analýzy sme sa zamerali aj na dotazník, ktorý vyplnili študenti na Slovensku vo veku 13-17 rokov, aby sme zistili ich názory a preferencie týkajúce sa štúdia v zahraničí. Odpovede ukazujú rozmanitú paletu záujmov a motivácií, ktoré sa však často kryjú v určitých spoločných bodoch.

Geografická a jazyková blízkosť

Mnohí študenti uprednostňujú krajiny blízke Slovensku, a to geograficky i jazykovo. Česká Republika je v tomto ohľade často spomínaná, hlavne kvôli minimálnej jazykovej bariére a blízkosti k slovenským hraniciam. To je zrozumiteľné, keďže pre Čechy a Slovákov platí zjednodušený režim pri uznávaní vzdelania, a krajiny majú aj podobné historické a kultúrne tradície. Vysoká úroveň niektorých českých univerzít je potvrdená v medzinárodných hodnoteniach, ako sú napríklad QS World University Rankings.

Kultúra a životný štýl

Odpovede ukazujú, že kultúra a životný štýl sú dôležitými faktormi pri výbere krajiny na štúdium. Francúzsko je známe svojím bohatým kultúrnym dedičstvom, vrátane múzeí, divadiel a hudobných festivalov, a teda priťahuje študentov, ktorí sa chcú kultúrne obohatiť. Španielsko je zase populárne pre svoju otvorenú a pohostinnú kultúru, ale aj pre dobre hodnotené univerzity, ako je Universidad de Barcelona alebo Universidad Autónoma de Madrid.

Kvalita vzdelania

Študenti vnímajú krajiny ako Fínsko, Veľká Británia a Nemecko ako synonyma vysokého štandardu vzdelania. Fínský vzdelávací systém je často hodnotený ako jeden z najlepších na svete v PISA testoch. Veľká Británia je domovom niekoľkých svetovo uznávaných univerzít, ako je Oxford alebo Cambridge. Nemecko sa môže pochváliť technickými univerzitami, ktoré sú na čele vo vede a výskume.

Osobný a profesionálny rozvoj

Mnohí študenti vnímajú Spojené štáty americké ako krajinu obrovských možností, čo nie je prekvapujúce. USA má najviac univerzít v rebríčkoch svetových škôl a je domovom Sillicon Valley, centra technologickej inovácie. Anglický jazyk je navyše globálnym lingua franca, a štúdium v anglojazyčnom prostredí je považované za veľkú výhodu na medzinárodnom pracovnom trhu.

Iné faktory

Niektoré odpovede reflektovali aj osobné motivácie, ako sú rodinné väzby alebo špecifické akademické záujmy. Zaujímavé je, že exotičtějšie destinácie ako Japonsko a Kórea sú vnímané ako miesta s kvalitným vzdelaním a jedinečnou kultúrou. Japonsko, napríklad, je známe svojím technologickým pokrokom a kultúrnymi ikonami ako je manga a anime.

Finančná stránka

Finančná náročnosť štúdia v zahraničí je implicitným faktorom, ktorý môže ovplyvniť konečné rozhodnutie študentov. V niektorých krajinách, ako sú napríklad severské štáty, je školné zdarma alebo veľmi nízke aj pre medzinárodných študentov, čo môže byť lákavé.

Zhrnuté a podčiarknuté, tieto odpovede a s nimi súvisiace fakty ukazujú, že motivácie mladých Slovákov sú rôznorodé, ale prelínajú sa v niekoľkých kľúčových aspektoch: geografickej a jazykovej blízkosti, kultúre, kvalite vzdelania, osobnom a profesionálnom rozvoji, a finančnej dostupnosti. Tieto zistenia môžu poslúžiť ako cenný vstup pre všetkých zainteresovaných strán: od politických tvorcov, cez akademické inštitúcie, až po samotných študentov a ich rodiny.

3 Opis parametrov

V našej práci využijeme 34 parametrov. Jednotlivé parametre sme zaradili do piatich hlavných skupín: Vzdelávanie, Ekonomické, Sociálne, Demografické a Ďalšie. V tejto kapitole predstavíme jednotlivé dáta a opíšeme, čo je oblasťou ich skúmania.

3.1 Vzdelávanie

Jedna z hlavných hypotéz znie: „Vzdelávanie má signifikantný vplyv pri výbere štúdia a to najmä ak študent pochádza z krajiny so slabšou úrovňou vzdelávacieho systému.“ Parametre sme vyberali na základe odpovedí študentov a vlastnej intuície a následne boli skonzultované a doplnené po konzultácii expertom - Miloslavom Bahnom zo Sociologického ústavu SAV.

3.1.1 Rebríček vzdelávania

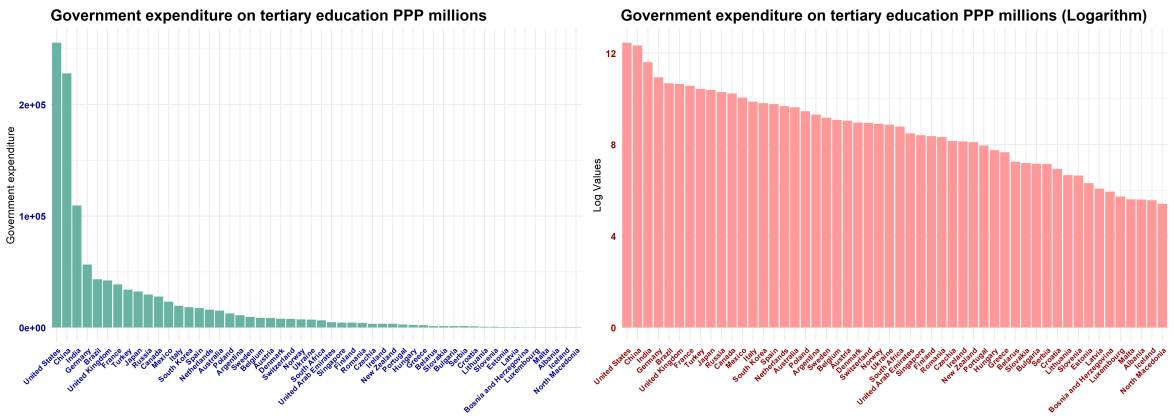
Ako parameter zachytávajúci kvalitu vzdelávania sme zvolili rebríček vzdelávania podľa krajín z portálu Wisevoter. Tento rebríček hodnotí úroveň vzdelávacieho systému jednotlivých krajín. Krajinám sú pridelené indexy z intervalu interval [0, 1], pričom 1 predstavuje najlepšie a 0 najhoršie hodnotenie (*zdroj:* [20]). Tento rebríček

bol vytvorený v roku 2022. Zmeny a vylepšenia vzdelávacieho systému nepovažujeme za proces, ktorý by sa odohrával rýchlo alebo by sa významne menil z roka na rok. Vzdelávací systém sa vyvíja postupne, pričom kladie dôraz na dlhodobú stabilitu a kontinuitu. To znamená, že zmeny v tomto sektore prebiehajú pomalým a premysleným tempom. Preto predpokladáme, že zmeny v rebríčku oproti rokom 2016 až 2019 nebudú signifikantné, a rebríček z roku 2022 považujeme za ich dobrú aproximáciu.

3.1.2 Vládne výdavky na vysokoškolské vzdelávanie ako % HDP a Vládne výdavky na vysokoškolské vzdelávanie,v PKS v miliónoch $

Ako parameter zachytávajúci rozvoj vzdelávania v danej krajine v spojitosti dostupných prostriedkov školstva sme použili dva parametre a to: Vládne výdavky na vysokoškolské vzdelávanie, ako % HDP a Vládne výdavky na vysokoškolské vzdelávanie, v miliónoch $ v parite kúpnej sily (*zdroj:* [14]). Pre konštrukciu oboch parametrov sme využili dostupné dáta za roky 2016 až 2019 a spravili sme priemer týchto hodnôt ako výsledné dáto pre danú krajinu.

V prípade druhého parametra na dáta použijeme transformáciu prirodzeným logaritmom, aby sme zabezpečili linearizáciu vzťahu, obrázok 1. Nemyslíme si, že by percepciu študenta ovplyvnili veľké rozdiely vo výdavkoch štátov. V praktickom svete študent vníma rozdiely hodnôt prirodzene podľa informácií o krajine a teda skôr lineárne.

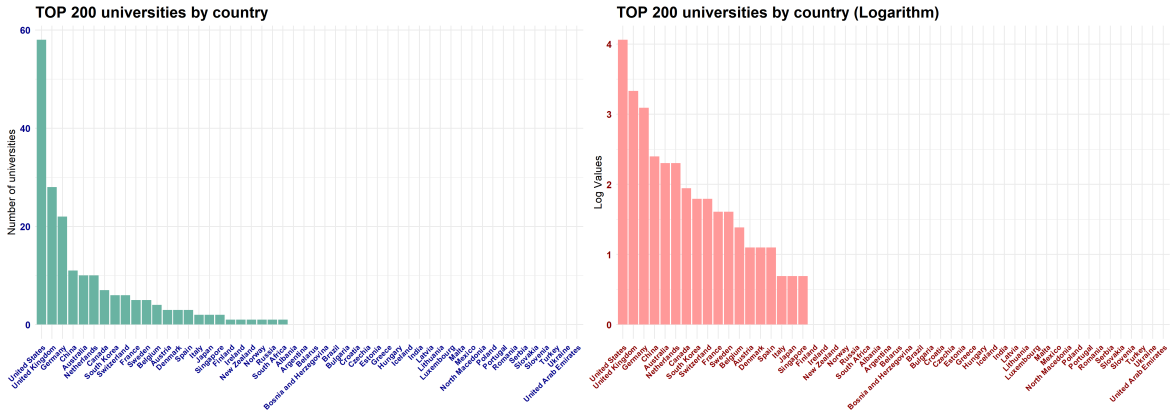


Obr. 1

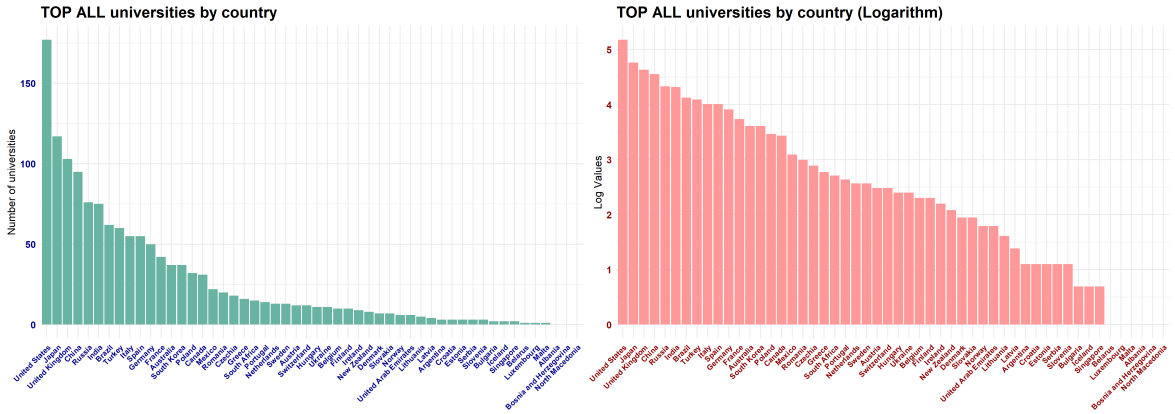
3.1.3 Svetový rebríček univerzít 2023

Kvalita vysokých škôl hrá významnú úlohu pri výbere krajiny pre štúdium. Ďalším dôležitým faktorom je počet vysokých škôl v krajine. Do väčších krajín s rozsiahlejším portfóliom študijných programov cestuje viac študentov ako do krajín, kde je výber študijných programov menší. Preto potrebujeme parametre, ktoré nám umožnia zohľadniť oba tieto aspekty. Ako podklad pre vytvorenie metrík sme použili Svetový rebríček univerzít portálu Times Higher Education (*zdroj:* [11]) z roku 2023, keďže sme predpokladali, že zmeny v rebríčku nebudú výrazne iné oproti rokom 2016 až 2019. Pre zostrojenie jednotlivých parametrov sme použili skóre Overall. Bližšie informácie o vyhodnocovaní rebríčka nájdete v zdroji [12].

Rozhodli sme sa vytvoriť 8 parametrov. TOP 200 rebríčky obsahujú počet univerzít v krajine a sumu bodov univerzít v prvých 200 univerzitách rebríčka. Tieto parametre selektujú krajiny s najlepšími univerzitami sveta. TOP 750 rebríčky obsahujú počet univerzít v krajine a sumu bodov univerzít v prvých 750 univerzitách rebríčka. Tieto parametre zachytávajú aspoň veľmi dobré univerzity sveta. TOP 1500 rebríčky obsahujú počet univerzít v krajine a sumu bodov univerzít v prvých 1500 univerzitách rebríčka. Tieto parametre zachytávajú aspoň dobré univerzity sveta. TOP ALL rebríčky obsahujú počet univerzít v krajine a sumu bodov univerzít všetkých univerzít rebríčka. Na každý z parametrov sme sa rozhodli aplikovať logaritmickú transformáciu, aby sme zabezpečili linearizáciu dát. Obrázok 2 a Obrázok 3.



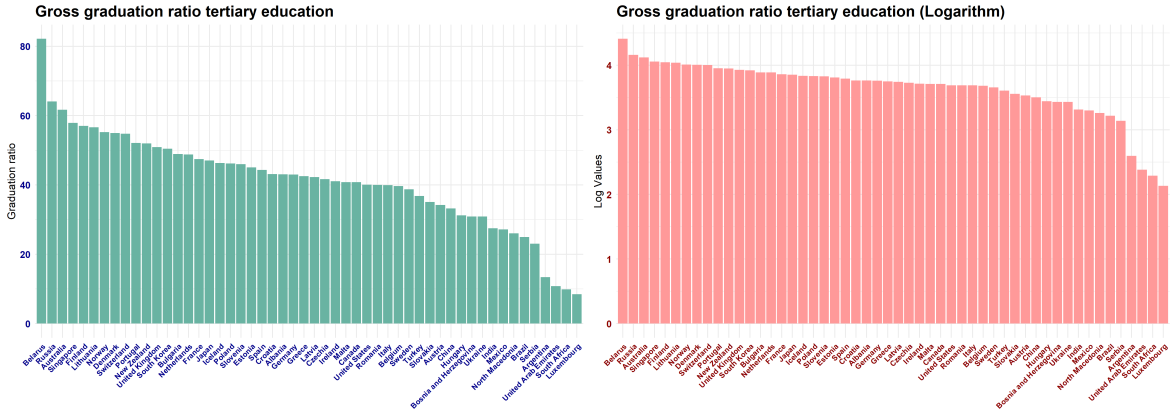
Obr. 2



Obr. 3

3.1.4 Hrubý pomer ukončenia vysokoškolského vzdelávania

Motiváciou študenta pre zahraničné štúdium môže byť aj pravdepodobnosť dokončenia štúdia na vysokej škole. Preto sme ako ďalší parameter vybrali Hrubý pomer ukončenia vysokoškolského štúdia, teda aké percento študentov, ktorí začnú študovať na vysokej škole, svoje štúdium dokončí. Vychádzali sme z databázy UNESCO (*zdroj:* [15]). Na vytvorenie parametru sme zobrali dáta pre krajiny vzorky z rokov 2016 až 2019 a spravili sme ich priemer. Výsledné hodnoty sú z intervalu interval $[0, 100]$ a reflektujú percentá hrubého pomeru dokončenia vysokoškolského štúdia.



Obr. 4

Ako vidieť na obrázku 4, jediným outlierom spomedzi krajín je Bielorusko. Keďže nepovažujeme, že by veľký rozdiel v pomere hrubého dokončenia vysokoškolského štúdia v Bielorusku mal signifikantný vplyv na prílev študentov do tejto krajiny, použijeme na dáta logaritmickú transformáciu pre odstránenie tohoto outlieru a zlinearizovanie dát.

3.1.5 Výdavky na vývoj a výskum ako percento z HDP

Možnosť uplatnenia sa v oblasti vedy a zapojenie do vedeckých projektov môže hrať významnú úlohu pri výbere krajiny na vysokoškolské štúdium. Predpokladáme, že možnosti zapojenia sa do výskumu alebo vývoja sú úzko spojené s financovaním štátu v tejto oblasti. Preto sme ako ďalší parameter vybrali výdavky na výskum a vývoj ako percento HDP. Vychádzali sme z databázy UNESCO (*zdroj:* [15]), z ktorej sme získali dáta z rokov 2016 až 2019. Výsledný parameter bol skonštruovaný ako priemer hodnôt z jednotlivých rokov pre krajiny.

3.2 Ekonomické parametre

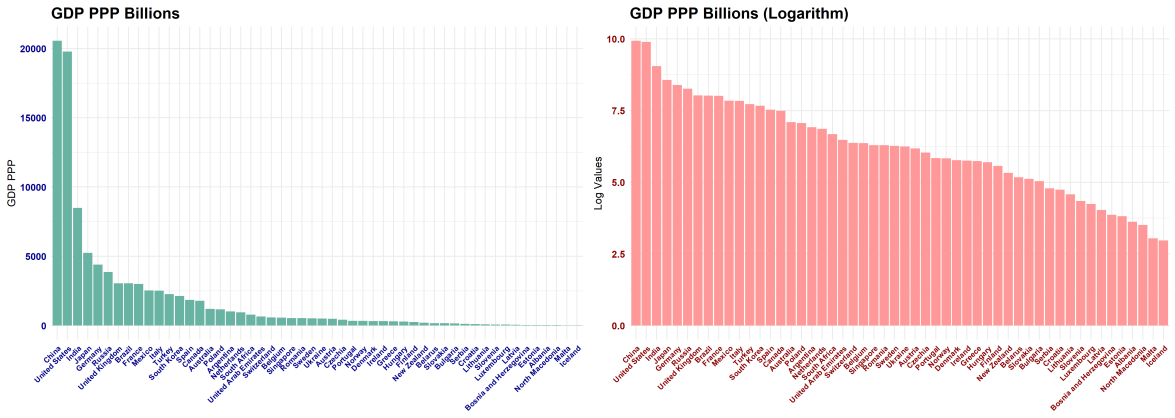
Dôležitým aspektom pri výbere štúdia je ekonomický stav krajiny. Predpokladáme, že študentou motivuje vidina stabilnej finančnej budúcnosti a veľký výber možností uplatnenia. Parametre sme vyberali na základe odpovedí študentov v dotazníku a vlastnej intuície. Následne boli doplnené po konzultácii expertom - Martinom Šusterom z Rady pre rozpočtovú zodpovednosť.

3.2.1 Parametre spojené s HDP

HDP je považované za jeden z najlepších indikátorov ekonomickej výkonnosti a stability krajiny. Hoci študenti uvažujúci o štúdiu v zahraničí s cieľom finančného zlepšenia životnej úrovne nemusia priamo sledovať HDP krajiny, ich vnímanie ekonomickej situácie je úzko späté s týmto ukazovateľom. Rozhodli sme sa preto použiť tri parametre spojené s HDP: HDP krajiny v parite kúpnej sily v medzinárodnom konštantnom dolári (*zdroj:* [17]), HDP na obyvateľa v parite kúpnej sily (*zdroj:* [17]) a medziročný rast HDP (*zdroj:* [16]), pričom všetky parametre predstavujú priemer za roky 2016 až 2019.

Po konzultácii s pánom Mgr. Martinom Šusterom, PhD., sme sa rozhodli v prvých dvoch parametroch nahradiť HDP Írska pomocou ukazovateľa Gross National Income, aby sme lepšie odrážali ekonomickú situáciu v krajine. Je to štandardný postup vzhľadom na to, že HDP Írska je umelo zvýšené pôsobením medzinárodných firiem využívajúcich daňové úľavy v krajine.

Pri parametroch HDP krajiny v parite kúpnej sily sme aplikovali logaritmickú transformáciu za účelom zlinearizovania dát, keďže percepcia študentov je skôr lineárna, viď obrázok 5.



Obr. 5

3.2.2 Priemerný príjem PPP

Jedným z relevantných ukazovateľov ekonomickej situácie v krajine je priemerný príjem (*zdroj:* [9]). Neexistuje celistvá databáza priemerných príjmov v krajine pre našu vzorku krajín. Preto sme tento parameter dopĺňali s dostupných informácii krajiny. Tento parameter vzniká prenásobením priemerných miezd v krajine indexom parity kúpnej sily.

3.2.3 Giniho index 2020

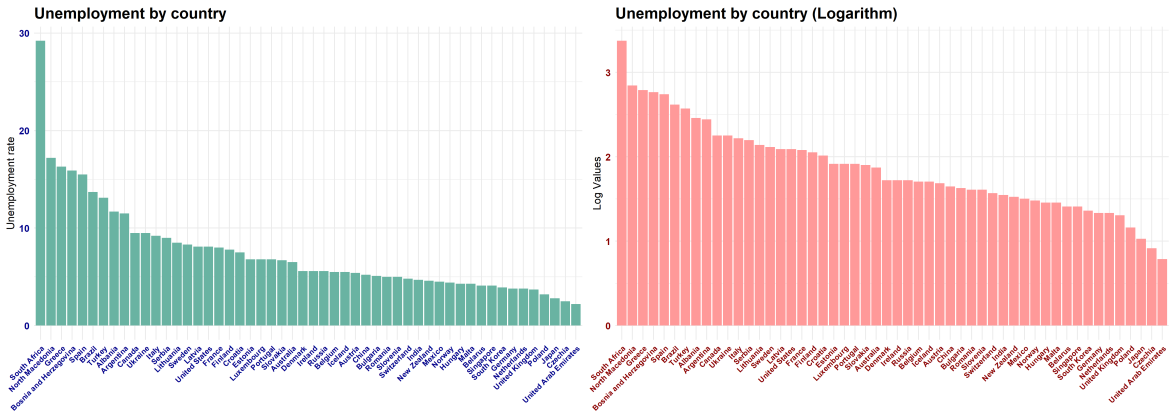
Výrazné rozdiely medzi bohatými a chudobnými vrstvami môžu vytvárať negatívny obraz štúdia v danej krajine. Preto sme začlenili Giniho index ako ďalší parameter do našej štúdie. Táto štatistika odhaduje celosvetové poradie Giniho indexu v roku 2020 (*zdroj:* [10]), diferencovaného podľa jednotlivých krajín. Giniho index meria rozsah príjmovej nerovnosti na škále od 0 (úplná rovnosť príjmov) po 1 (úplná nerovnosť).

3.2.4 Zamestnanosť a nezamestnanosť

Ukazovatele zamestnanosti a nezamestnanosti sú kľúčové pre posúdenie ekonomickej vitality krajiny a majú priamy vplyv na kvalitu života jej obyvateľov. Stabilná a prosperujúca ekonomika je schopná generovať nové pracovné miesta, čo vedie k poklesu nezamestnanosti. Vzhľadom na to, že niektoré krajiny môžu nezamestnanosť umelo redukovať, relevantnejším ukazovateľom sa stáva miera zamestnanosti. Keďže spoľahlivé dáta pre všetky krajiny vzorky sme nenašli, rozhodli sme sa použiť údaje o zamestnanosti pre krajiny OECD a údaje o nezamestnanosti pre celú vzorku krajín.

Údaje o zamestnanosti sme čerpali z OECD za obdobie od 4. kvartálu 2021 do 3. kvartálu 2023 (*zdroj:* [8]). Tieto údaje sú postačujúce, keďže významné zmeny v zamestnanosti sú dlhodobého charakteru. Výsledný parameter predstavuje priemer hodnôt zamestnanosti za toto obdobie.

Pri nezamestnanosti sme využili údaje od World Population Review (*zdroj:* [19]). Pre väčšinu krajín vzorky sme použili údaje z roku 2020. Pre krajiny, pre ktoré boli dostupné len staršie údaje, sme vzali najstarší dostupný rok z databázy. Aby sme dosiahli linearizáciu dát, rozhodli sme sa pre logaritmickú transformáciu, ako je znázornené na obrázku 6.



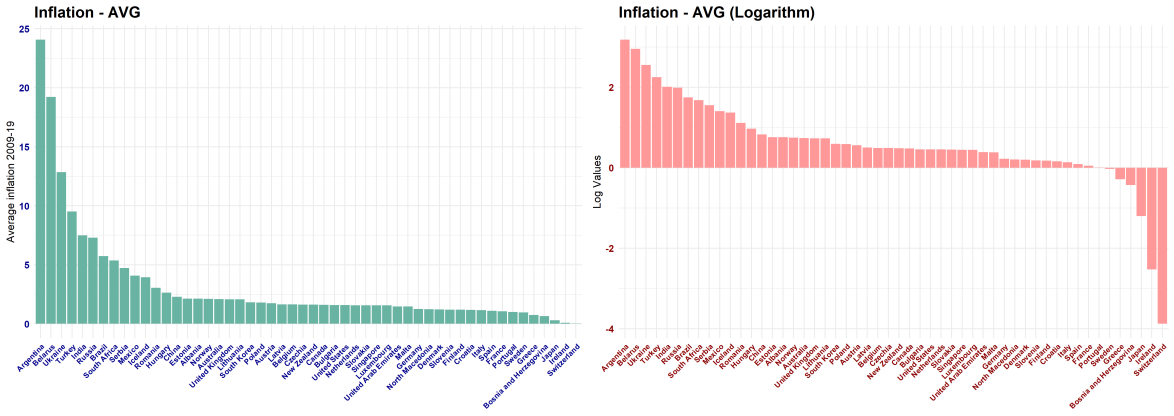
Obr. 6

3.2.5 Inflácia

Inflácia je dôležitým makroekonomickým ukazovateľom, ktorý však reflektuje krátkodobé ekonomické podmienky. Dlhotrvajúca vysoká inflácia môže signalizovať

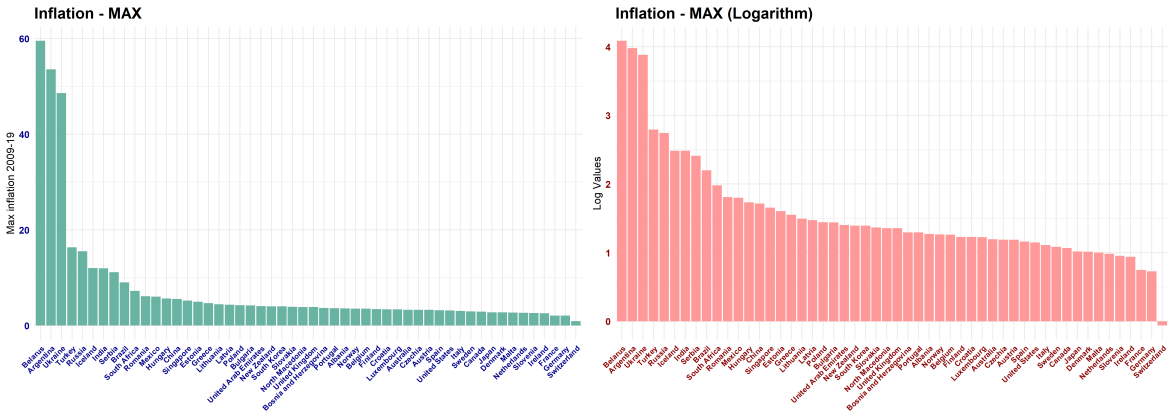
ekonomickú nestabilitu krajiny alebo zvýšené riziko. Ako základ sme použili vývoj inflácie v sledovaných krajinách za obdobie 2009 až 2019 (*zdroj:* [18]).

Vytvorili sme tri rozličné parametre. Prvým je celkový priemer inflácie v danom období, čím sme eliminovali význam krátkodobých fluktuácií inflácie. Vyššie hodnoty tohto parametra môžu indikovať dlhodobú ekonomickú nestabilitu. Na tento parameter sme neaplikovali logaritmickú transformáciu, aby sme zachovali väčšiu váhu vysokým hodnotám priemernej inflácie, viď obrázok 7.



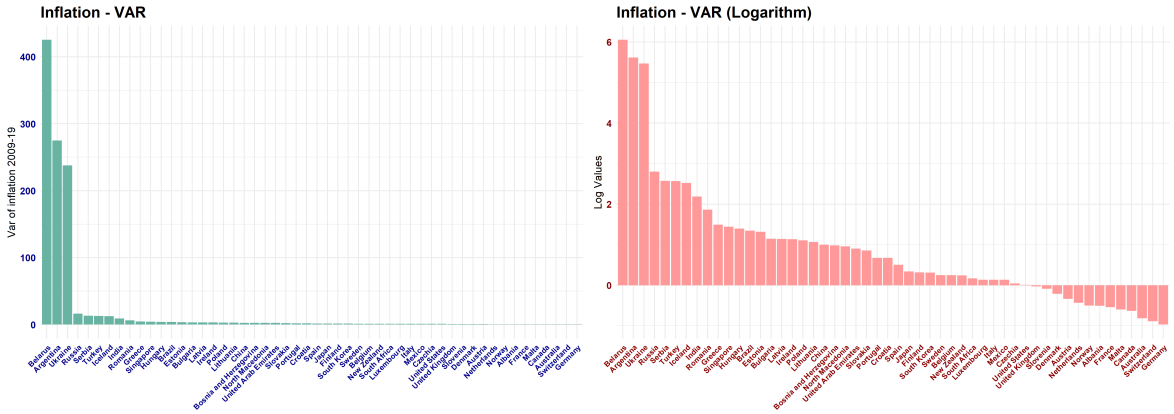
Obr. 7

Druhým parametrom je maximálna hodnota inflácie v sledovanom období, ktorá poskytuje prehľad o extrémnych hodnotách ovplyvňujúcich ekonomiku. Vysoká inflácia môže znižovať kupnú silu meny a ovplyvňovať úspory a investície. Pre študentov uvažujúcich o štúdiu v zahraničí by maximálna inflácia mohla ovplyvniť ich finančné rozhodnutia a náklady na život. Pre zlinearizovanie dát sme zvolili logaritmickú transformáciu, ako ukazuje obrázok 8.



Obr. 8

Tretí parameter, rozptyl inflácie, slúži ako odhad volatility inflácie v jednotlivých krajinách. Krajiny s nižšou ekonomickou stabilitou budú mať vyššie hodnoty rozptylu ako stabilnejšie ekonomiky. Aj tu sme pre linearizáciu údajov uplatnili logaritmickú transformáciu, obrázok 9.



Obr. 9

3.3 Sociálne parametre

Sociálne parametre sú pre študentov, ktorí plánujú študovať v zahraničí, nesmierne dôležité, pretože ovplyvňujú ich schopnosť prispôsobiť sa novej kultúre, integrovať sa do spoločnosti a cítiť sa bezpečne. Tieto faktory, vrátane politické stability, kultúrnej diverzity, sociálneho systému, úrovne korupcie a životného prostredia, majú priamy vplyv na ich vzdelávaciu skúsenosť a celkovú spokojnosť. Správne pochopenie týchto aspektov im umožňuje lepšie sa pripraviť na nové prostredie a efektívne sa v ňom zorientovať.

3.3.1 Politická stabilita a absencia násilia/terorizmu odhad

Politická stabilita je jeden z fascinujúcich sociálnych parametrov, ktoré by mohli ovplyvniť rozhodovanie študentov pri výbere krajiny. Avšak, meranie tohto parametra môže byť náročné. V databáze Svetovej banky (*zdroj:* [21]) sme našli parameter Politická stabilita a absencia násilia/terorizmu. Konečný parameter bol vytvorený ako priemer hodnôt z rokov 2016 až 2019.

3.3.2 Dĺžka života

Priemerná dĺžka života v krajine odráža životnú úroveň obyvateľov a je tiež úzko spojená so zdravotným systémom a úrovňou krajiny. Tento parameter by preto mohol ovplyvniť rozhodnutie študentov pri výbere krajiny pre štúdium. Dáta sme čerpali z databázy Svetovej banky (*zdroj:* [22]), pričom parameter bol vytvorený ako priemer dĺžky života v krajine v rokoch 2016 až 2019.

3.3.3 Rebríček korupcie

Pri výbere krajiny na štúdium by mali študenti zvážiť úroveň korupcie, keďže môže negatívne ovplyvniť kvalitu a medzinárodné uznanie vzdelávania, spôsobovať komplikácie v administratívnych procesoch ako získavanie víz a študentských povolení, a môže zvyšovať životné náklady. Vysoká úroveň korupcie môže tiež indikovať nestabilitu a bezpečnostné riziká v krajine, čo by mohlo ovplyvniť celkovú študentskú skúsenosť a bezpečnosť. Pre našu prácu sme sa rozhodli použiť dáta Transparency International z roku 2018 (*zdroj:* [23]).

3.3.4 Jazykové vzdialenosti krajín na základe Levensteinových vzdialeností

Jazykové vzdialenosti medzi krajinami predstavujú pre študentov dôležitý parameter pri výbere krajiny pre ich ďalšie štúdium. Študentov prirodzene lákajú krajiny, v ktorých sa môžu dorozumieť materinským jazykom, ale taktiež krajiny, kde sa dohovoria v jazykoch, ktoré ovládajú a vedia plynule používať. Napríklad v prípade slovanských krajín hrá veľkú rolu aj schopnosť porozumieť týmto jazykom bez hlbokých znalostí.

Jazykové vzdialenosti sú však veľmi komplikovanou premennou, a to nielen z matematického hľadiska, ale aj z pohľadu zhromažďovania dát z dostatočného počtu krajín pre niektorý z modelov v našej práci. Z tohto dôvodu tento parameter zaradíme do Globálneho modelu prvej skupiny dát v samostatnej sekcii, kde špecificky upravíme celý dátový rámec.

Keďže na internete bežne nie sú dostupné hotové jazykové vzdialenosti medzi krajinami, odvodíme vzorec, podľa ktorého budeme počítať jazykové vzdialenosti v našej

práci. Opierať sa pritom budeme o dve tabuľky. Prvá tabuľka reprezentuje percentuálne zastúpenie občanov hovoriacich konkrétnymi jazykmi v jednotlivých krajinách [5]. Tu narazíme na prvé obmedzenie v podobe existencie takýchto dát len pre 26 európskych krajín. Druhou nevyhnutnou tabuľkou sú Levensteinove vzdialenosti indoeurópskych jazykov, kde koeficient pozície i, j, kde $i \neq j$, predstavuje vzdialenosť medzi jazykmi i a j. Hodnota tohto koeficientu je z intervalu $[0, 1]$, pričom 0 znamená identické jazyky a 1 maximálne rozdielne. Zvyšujúci koeficient značí väčšiu vzdialenosť dvoch jazykov, a teda zväčšujúcu sa rozdielnosť jazykov. Metodiku výpočtu týchto vzdialeností nájdete v zdroji [6]. Pre ďalší priebeh výpočtu jazykových vzdialeností bolo potrebné upraviť tieto dáta, vybrať len jazyky, ktoré sa nachádzajú v tabuľke zdroja [5], a doplniť chýbajúce. V prípade chýbajúcich dát sme postupovali nasledovne: Ak boli dva jazyky veľmi rozdielne a pochádzali z rozličných jazykových rodín, priradili sme danému jazyku maximálnu hodnotu nachádzajúcu sa v našej tabuľke. V prípade podobných jazykov, teda jazykov z rovnakej jazykovej rodiny, sme sa snažili odhadnúť veľkosť koeficienta na základe dát z iných krajín.

Nech A je matica rozmerov 26×38, kde i-tý riadok zodpovedá hodnotám pre i-tu krajinu, pre $i = 1, \ldots, 26$. Stĺpec j, pre $j = 1, \ldots, 38$, zodpovedá j-temu jazyku. Hodnota na pozícii (i, j) predstavuje percentuálne zastúpenie obyvateľstva hovoriaceho jazykom j v krajine i.

Nech B je matica Levensteinových vzdialeností medzi jazykmi, kde rozmer matice B je 38×38 a prvky matice $B_{m,n}$ predstavujú Levensteinovu vzdialenosť medzi jazykom m a jazykom n, pre $m, n = 1, \ldots, 38$.

Potom maticu jazykových vzdialeností J rozmeru 26×26 dostaneme ako:

$$\underset{26\times26}{\mathrm{J}} = \underset{26\times38}{A} \times \underset{38\times38}{B} + \underset{38\times26}{A^T} \tag{1}$$

kde A^T je transponovaná matica matice A.

Pre krajiny s vysokou pravdepodobnosťou, že dvaja náhodne vybraní občania sa dorozumejú, bude tento koeficient blízky 1. V prípade krajín s nízkou pravdepodobnosťou dorozumenia medzi dvoma náhodne vybranými občanmi bude tento koeficient výrazne odlišný od 1.

Nevýhodou parametra jazykových vzdialeností je množstvo krajín, pre ktoré vieme

tento parameter vytvoriť. Konkrétne sme jazykové vzdialenosti vytvorili pre krajiny: Belgicko, Bulharsko, Česko, Dánsko, Estónsko, Fínsko, Francúzsko, Grécko, Holandsko, Írsko, Litva, Lotyšsko, Luxembursko, Maďarsko, Malta, Nemecko, Poľsko, Portugalsko, Rakúsko, Rumunsko, Slovensko, Slovinsko, Spojené kráľovstvo, Španielsko, Švédsko.

3.3.5 Etická, jazyková a náboženská frakcionalizácia

Pri výbere krajiny na štúdium by mohla zohrať úlohu etnická, jazyková alebo náboženská rozmanitosť krajiny. Takéto údaje možno nájsť na Wikipédii (*zdroj:* [24]), ktoré poskytujú prehľad z roku 2003. Vysoká miera rozmanitosti môže obohatiť vzdelávací zážitok prostredníctvom expozície bohatej kultúrnej dynamiky a rôznorodých perspektív. Na druhej strane, krajina s vysokou mierou rozmanitosti môže byť sprevádzaná sociálnymi napätiami, čo môže ovplyvniť koheziu a bezpečnosť, a to je dôležité brať do úvahy pri rozhodovaní o študijnom pobyte v zahraničí.

3.3.6 Index lepšieho života (Better Life Index)

Better Life Index je index vytvorený na porovnanie krajín OECD v jedenástich témach (*zdroj:* [7]). Slúži ako pomocník pri rozhodovacom procese pre ľudí, ktorí chcú migrovať v rámci krajín OECD. Obsahom Better Life Indexu je viacero zaujímavých parametrov. My sme sa však rozhodli vybrať len niekoľko z nich, ktoré by mohli mať signifikantný vplyv na migráciu študentov: náklady na bývanie, počet izieb na osobu, zamestnanosť, znečistenie vzduchu, kvalita vody, životná spokojnosť, pocit bezpečia pri nočnej prechádzke osamote a miera vrážd.

Pretože ide o zaujímavé a ťažko merateľné parametre a máme dáta len pre krajiny OECD, rozhodli sme sa vytvoriť druhú sadu dát na bližšiu analýzu, ktorá bude zameraná len na krajiny OECD a zároveň bude obsahovať aj dáta z Better Life Indexu.

3.3.7 Minulé toky študentov

Pre študentov môže byť kľúčové, či poznajú niekoho, kto študuje alebo študoval v danej krajine. Taktiež minulé toky môžu odrážať popularitu tejto destinácie v krajine odlevu. Preto použijeme minulé toky študentov ako ďalší parameter. Vychádzať budeme

z dát odlevu študentov, ale namiesto rokov 2016 až 2019 zoberieme interval rokov 2012 až 2015 a vypočítame priemer tokov za tieto roky.

Na parameter minulých tokoch študentov aplikujeme logaritmickú transformáciu. Dôvodom je kategorické vnímanie študentov, teda študent vníma krajiny kategoricky ako krajiny kde chodí málo, stredne a veľa študentov s jeho štátnou príslušnosťou.

3.4 Demografické parametre

Často motivácia študentov pre vysokoškolské štúdium vychádza z demografických parametrov. Mnohí študenti hľadajú možnosti zahraničného štúdia, ktoré sú bližšie k domovu, kvôli lepším logistickým možnostiam a tiež kvôli možnosti častejšieho návratu domov. Na druhej strane, veľké krajiny môžu byť pre študentov obrovským lákadlom, kvôli perspektíve sociálneho rozvoja a lepším podmienkam pre „prežitie“ obdobia vysokoškolského štúdia.

3.4.1 Susedné štáty

Pri rozhodovaní o zahraničnom štúdiu môže parameter susedných krajín hrať významnú úlohu. Susedné štáty často zdieľajú kultúrne a jazykové podobnosti, čo môže študentom uľahčiť adaptáciu a zlepšiť ich vzdelávací zážitok. Blízkosť susedných krajín tiež zvyčajne znamená lepšie logistické možnosti a jednoduchšie cestovanie, čo je výhodné pre časté návštevy domova alebo pre prípad núdzových situácií. Okrem toho regionálne dohody môžu poskytovať štipendiá alebo zjednodušené vízové režimy.

V našej práci využijeme dáta z diplomovej práce Filipa Šramka (*zdroj:* [13]). Tieto dáta ale sú len pre krajiny OECD. Preto sme pre ostatné krajiny vzorky manuálne doplnili chýbajúce dáta pre jednotlivé krajiny. Brali sme do úvahy len ak štáty majú suchozemskú hranicu, nie hranicu v námornom priestore.

3.4.2 Populácia

Veľkosť krajiny môže vplývať nepriamo na výber krajiny štúdia pre študentov. Pri väčších krajinách je pravdepodobnosť, že krajina má rozmanitú možnosť štúdijných programov. Taktiež niektorí študenti chcú študovať vo väčších mestách. Preto sme sa

rozhodli ako ďalší parameter vybrať veľkosť populácie krajiny. Keďže populácia sa veľmi nemení v priebehu rokov, zobrali sme dáta z roku 2023 (*zdroj:* [25]). Percepcia študentov pri zvažovaní tohoto parametra bude mať skôr lineárny charakter. Pre zlinearizovanie dát sme na dáta aplikovali logaritmickú transformáciu.

3.5 Vzdialenosť dvoch krajín - Vážená geografická vzdialenosť a vážená časova vzdialenosť cesty autom

Z odpovedí, ktoré sme zozbierali od študentov stredných škôl, vyplýva, že vzdialenosť študijného miesta od ich domovskej krajiny hrá dôležitú rolu pri výbere vysokej školy a krajiny, v ktorej sa škola nachádza. Tu však narazíme na prvý veľký problém našej práce: Čo vlastne znamená vzdialenosť dvoch krajín a ako ju merať? Túto problematiku skúmalo viacero vedeckých prác a článkov, no v žiadnom sa nenájde odpoveď na presný spôsob merania vzdialenosti medzi krajinami. Napríklad v práci (vložte presmerovanie na prácu Baláža + nájdite ešte nejakú prácu s iným spôsobom merania) autor používa na meranie vzdialenosti medzi krajinami vzdialenosť medzi hlavnými mestami. Pri tejto metóde však narazíme na problém, ak sú hlavné mestá dvoch krajín príliš blízko alebo príliš ďaleko od seba. Napríklad Bratislava je od Viedne vzdialená 70 km a približne 50 minút cesty autom. To by naznačovalo, že Slovensko je veľmi blízko Rakúska, bližšie než k akejkoľvek inej susednej krajine. Keď však overujeme túto hypotézu z pohľadu náhodných ľudí žijúcich na Slovensku, zistíme, že pre obyvateľov Bratislavy a okolia, resp. vybraných častí Západného Slovenska, sa táto hypotéza javí ako pravdivá. Na druhej strane, z pohľadu človeka žijúceho v Košiciach sa ako najbližšia krajina javí Maďarsko, zatiaľ čo pre obyvateľa Žiliny sú to Poľsko a Česká republika.

Pri tvorbe výpočtu budeme zohľadňovať len vzdialenosť autom a časovú vzdialenosť autom. Ak chceme získať dobrý odhad vzdialenosti medzi krajinami, mali by sme nájsť metódu, ktorá zohľadňuje aj názory ľudí žijúcich v častiach krajiny mimo hlavných miest. Dôležitá je tiež hustota osídlenia danej oblasti, resp. populácia žijúca v tejto oblasti. Napríklad Rakúsko je najbližšie pre 430 000 obyvateľov Bratislavy, pre 240 000 obyvateľov Košíc je najbližšou krajinou Maďarsko a pre 81 000 obyvateľov Žiliny sú najbližšími krajinami Poľsko a Česká republika. Ak by sme teda vážili najbližšiu

krajinu pre dané mesto podľa počtu obyvateľov, vyšlo by nám znovu Rakúsko.

Avšak, keďže hľadáme najbližšiu možnú krajinu pre všetkých ľudí žijúcich v každej z vybraných oblastí Slovenska, potrebujeme vo výpočtoch zohľadniť vzdialenosť do každej z krajín z každej našej oblasti. Vo všeobecnosti by sme vybrali n oblastí so signifikantným počtom obyvateľov a merali by sme vzdialenosť do jednotlivých krajín. Pre našu zjednodušenú interpretáciu myšlienky použijeme opäť pohľady obyvateľov jednotlivých miest zo vzorky a vzdialenosti budeme merať len v podobe dlhej, stredne dlhej a krátkej cesty. Vzdialenosti z vybraných miest do destinácii vyzerajú nasledovne:

- Rakúsko: občan Bratislavy – krátka cesta, občan Žiliny – stredne dlhá cesta, občan Košíc – dlhá cesta
- Maďarsko: občan Bratislavy – krátka cesta, občan Žiliny – dlhá cesta, občan Košíc – krátka cesta
- Česko: občan Bratislavy – krátka cesta, občan Žiliny – krátka cesta, občan Košíc – dlhá cesta

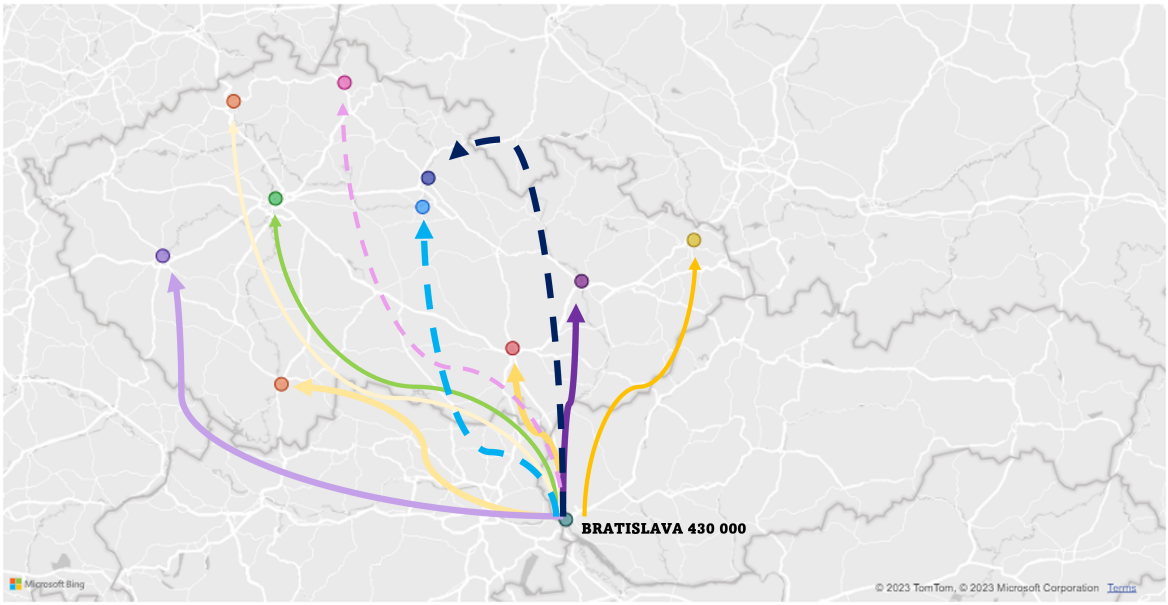
Tento spôsob merania však neprináša úplne jednoznačné výsledky. Keby sme pridali ďalšie slovenské mestá a ostatné susedné krajiny, výsledky by sa ešte viac premiešali.

Poslednou otázkou zostáva, ako vybrať miesto alebo oblasť v destinácii (krajine, kam cestujeme), od ktorej by sme mali merať jednotlivé vzdialenosti. Odpoveď máme: stačí nám použiť rovnakú logiku ako v prípade začiatočnej krajiny. Vyberieme niekoľko oblastí, kde žije najviac obyvateľov destinácie, a následne vytvoríme permutácie vzdialeností medzi oblasťami začiatočnej krajiny a oblasťami destinácie. Na základe predchádzajúcich úvah sme odvodili podmienky, ktoré musí naša metóda pri výpočte vzdialeností zohľadniť, aby sme mohli spraviť dobrú aproximáciu vzdialeností medzi dvoma krajinami:

1. Vzdialenosť oblastí destinácie od vybraných oblastí začiatočnej krajiny.
2. Počet obyvateľov žijúcich v jednotlivých oblastiach začiatočnej krajiny.
3. Vzdialenosť autom, zahŕňajúca časovú aj diaľkovú vzdialenosť.

3.5.1 Návrh výpočtu

Ako dôležité oblasti začiatočnej krajiny budeme považovať 10 najväčších miest danej krajiny. Označme mesto v krajine i ako c_{ik1}. Pre každé z týchto miest budeme



Obr. 10: Vzdialenosti z Bratislavy do Českej Republiky

potrebovať získať počet obyvateľov mesta, označených ako váha mesta $w_{ik1}^{(1)}$. Rovnako pre destináciu budeme potrebovať 10 najväčších miest, avšak v tomto prípade bez počtu obyvateľov. Označme mesto v krajine j ako c_{ik2}. Váhu miest destinácie označíme ako $w_{ik}^{(2)} = \frac{1}{10}$. Prvým krokom bude účinným spôsobom zmerať vzdialenosti medzi každou dvojicou miest začiatočnej krajiny a destinácie. To znamená, že pre dvojicu krajín A a B potrebujeme zmerať 100 rôznych vzdialeností. Nech začiatočná krajina je označená ako N_i pre $i = 1, 2, ..., 37$ a destinácia ako N_j, kde $j = 1, 2, ..., 52$. Vzdialenosť medzi týmito krajinami označíme ako funkciu $d(N_i, N_j)$. Vzdialenosť medzi krajinami $d(N_i, N_j)$ vypočítame nasledovne:

$$d_l(N_i, N_j) = \sum_{k_1=1}^{10} \sum_{k_2=1}^{10} \frac{(w_{ik_1}^{(1)} w_{ik_2}^{(2)} d_l(c_{ik_1}, c_{ik_2}))}{\sum_{k_1=1}^{10} w_{ik_1}^{(1)}} \tag{2}$$

A pre časové vzdialenosti:

$$d_t(N_i, N_j) = \sum_{k_1=1}^{10} \sum_{k_2=1}^{10} \frac{(w_{ik_1}^{(1)} w_{ik_2}^{(2)} d_t(c_{ik_1}, c_{ik_2}))}{\sum_{k_1=1}^{10} w_{ik_1}^{(1)}} \tag{3}$$

3.5.2 Implementácia a praktická aplikácia odvodeného vzorca

Výpočet budeme aplikovať na celkovo 52 rôznych krajín. Použijeme segmentáciu krajín z časti (nejaká časť Predstavenie krajín), pričom začiatočné krajiny označíme ako ORIGINS a destinácie ako DESTINATIONS. Pre každú krajinu zo zoznamu budeme potrebovať 10 najväčších miest a pre krajiny z časti ORIGINS budeme potrebovať aj populáciu jednotlivých miest. Pre zber dát sme využili UI jazykový model ChatGPT4, ktorého výstupom bol súbor vo formáte csv obsahujúci tri stĺpce: Country, City, Population. Posledným krokom bolo oddelenie krajín, do ktorých sa nedá dostať autom, od tých, kde je cesta autom možná. Následne sme pokračovali len s krajinami, medzi ktorými je automobilová cesta možná. Následne sme z tohto súboru skriptom v programovacom jazyku R, nazvaným city_pairs, vytvorili všetky permutácie dvojíc miest s dodržaním nasledujúcich podmienok:

- Dvojice sú jedinečné.
- Dvojice miest nie sú z tej istej krajiny.
- Dvojica c_{ik}, c_{jk} a c_{jk}, c_{ik} je považovaná za totožnú (v súbore vystupuje iba raz).

Následne sme si pripravili skript RawCsvWriter, v ktorom sme vypočítali vzdialenosti. Bolo to možné vďaka interaktívnemu napojeniu Google Distance API do prostredia R prostredníctvom knižnice gmapsdistance. Podrobná dokumentácia k gmapsdistance a príklady použitia v R sú k dispozícii v prílohe. Vďaka predošlým úpravám sme dokázali znížiť celkový počet volaní API z 265200 na 77400. Výsledným datasetom bol súbor csv so stĺpcami origin_country, origin_city, population, destination_country, destination_city, distance_km, distance_time. Posledným krokom bolo vypočítať jednotlivé vzdialenosti medzi krajinami pomocou vzorca uvedeného v časti (tej a tej). Pre lepšiu interpretáciu a prehľadnosť výsledkov sme dáta uložili vo forme matice, kde stĺpce prislúchajú jednotlivým DESTINATIONS a riadky ORIGINS. Všetky použité skripty a ich podrobné popisy nájdete v prílohe tohto dokumentu.

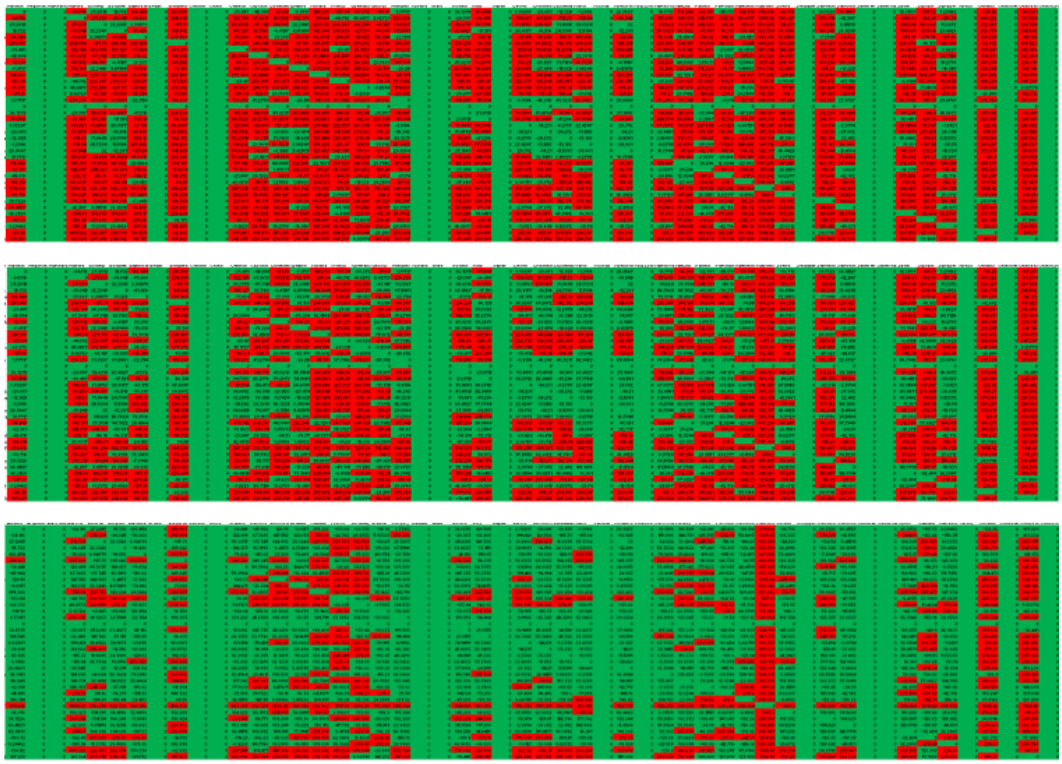
3.5.3 Porovnanie vzdialeností hlavných miest

V rámci riešenia problému vzdialenosti dvoch krajín sme sa zamerali na vylepšenie výpočtu vzdialeností medzi krajinami, pretože sme identifikovali, že vzdialenosť medzi hlavnými mestami (ďalej označovaná ako DoC – Distance of Capitals) nie je dostatočne presným ukazovateľom. V poslednej fáze tejto časti sme porovnali výsledky nášho výpočtu s tradičným meraním vzdialeností pomocou metódy DoC. Toto porovnanie je podrobne ilustrované obrázkami (čísla obrázkov).

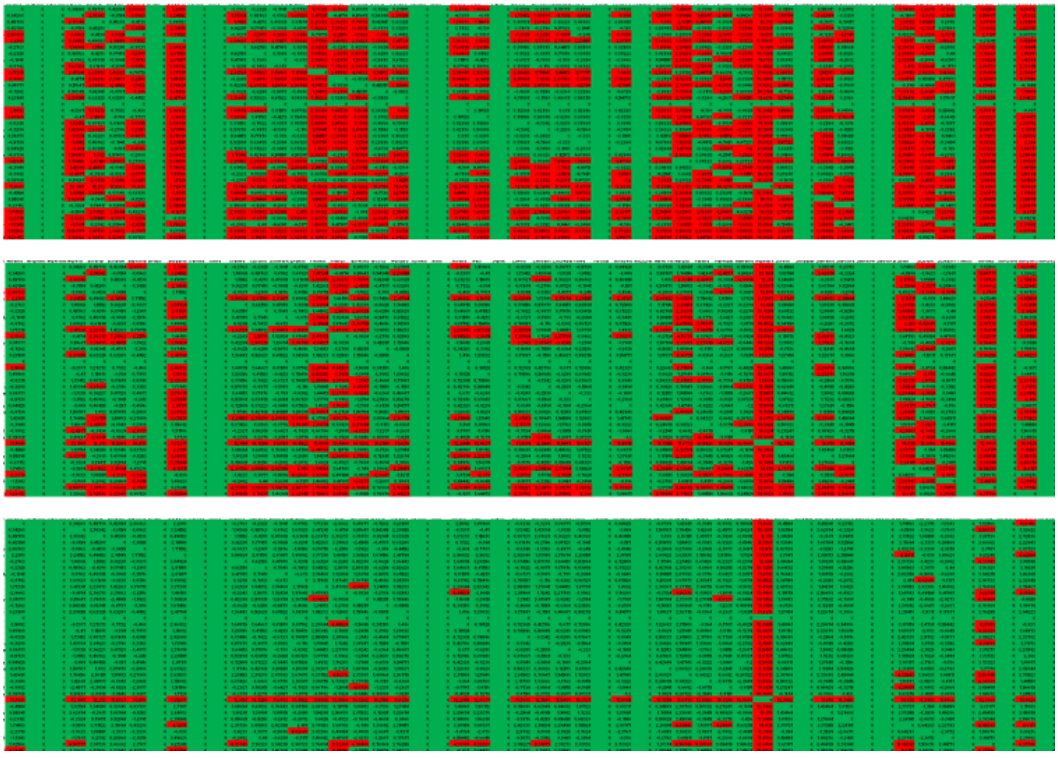
Pri vytváraní finálnej matice vzdialeností sme zaviedli zjednodušenie (ďalej nazývané AVG): ak sa krajina A objavuje ako začiatočná (ORIGINS) a krajina B ako cieľová (DESTINATIONS), a naopak, potom vzdialenosť medzi týmito krajinami sme vypočítali ako priemer hodnôt $d(N_i, N_j)$ a $d(N_j, N_i)$. Tento výpočet sme vykonali pomocou skriptu v programovacom jazyku R s názvom AVG. Matematický vzorec tohto priemerného výpočtu je:

$$\text{Priemerná vzdialenosť} = \frac{d(N_i, N_j) + d(N_j, N_i))}{2} \tag{4}$$

Porovnanie metódy DoC a AVG je ilustrované obrázkami (doplniť, ktorými). Na pozorovanie sme si zvolili hranice rozdielov 50 km, 100 km a 200 km v prípade vzdialeností a v prípade časových vzdialeností sme sledovali rozdiely 1 h, 2 h a 4 h. Vyšli nám rozdiely medzi metódou DoC a AVG, čo poukazuje na zlepšenie odhadu vzdialeností medzi krajinami, obrázoky 11 a 12.



Obr. 11: Porovnanie AVG a DoC - Rozdiel 50km, 100km, 200km



Obr. 12: Porovnanie AVG a DoC - Rozdiel 1h, 2h, 4h

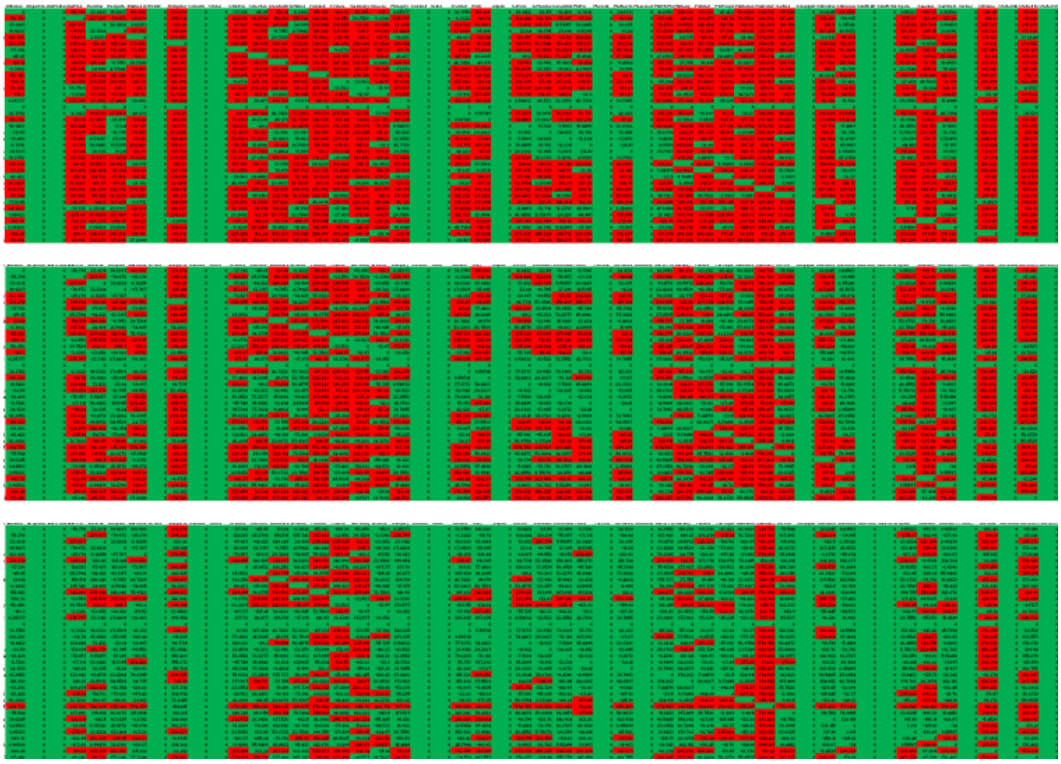
Okrem toho sme sa rozhodli vykonať aj druhý prístup (ďalej nazývaný WAVG), kde sme v porovnaní s AVG navyše vážili vzdialenosti z každého pohľadu podľa počtu

obyvateľov začiatočnej krajiny v 10 najväčších mestách danej krajiny. Vzorec takéhoto postupu vyzerá nasledovne:

$$\text{Vážená priemerná vzdialenosť} = \frac{\sum(w_{ik}^{(1)} d(N_i, N_j)) + \sum(w_{jk}^{(1)} d(N_j, N_i))}{\sum(w_{ik}^{(1)}) + \sum(w_{jk}^{(1)})} \tag{5}$$

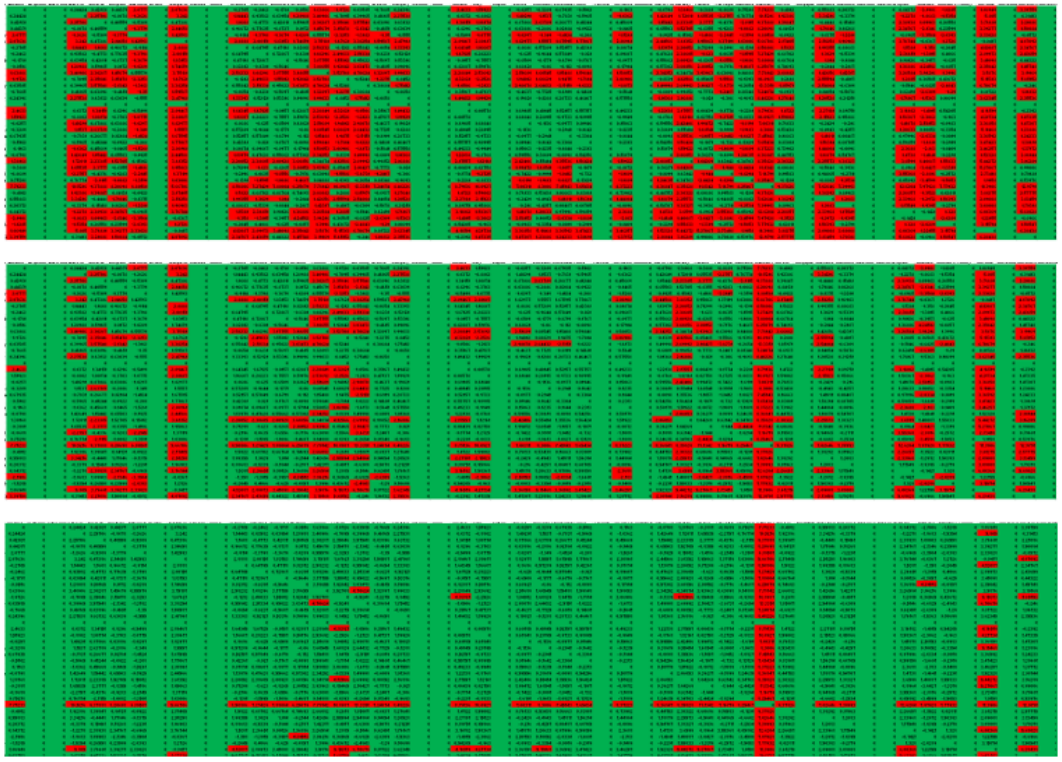
Tento výpočet bol realizovaný skriptom WAVG v R. Našou hypotézou bolo, že tento prístup nám umožní získať presnejšie a reprezentatívnejšie výsledky pre vzdialenosti medzi krajinami, čím by sme výrazne zlepšili celkovú spoľahlivosť a presnosť našich výsledkov. Porovnanie metódy DoC a WAVG je ilustrované obrázkami 13 a 14.

Tak isto, ako pri porovnávaní DoC a AVG, sme pozorovali rozdiely 50 km, 100 km, 200 km a v prípade časových vzdialeností 1 h, 2 h a 4 h. Opäť rozdiely vyšli ľahko pozorovateľné a taktiež možno konštatovať, že aj metóda WAVG prináša lepšie odhady vzdialeností medzi krajinami.



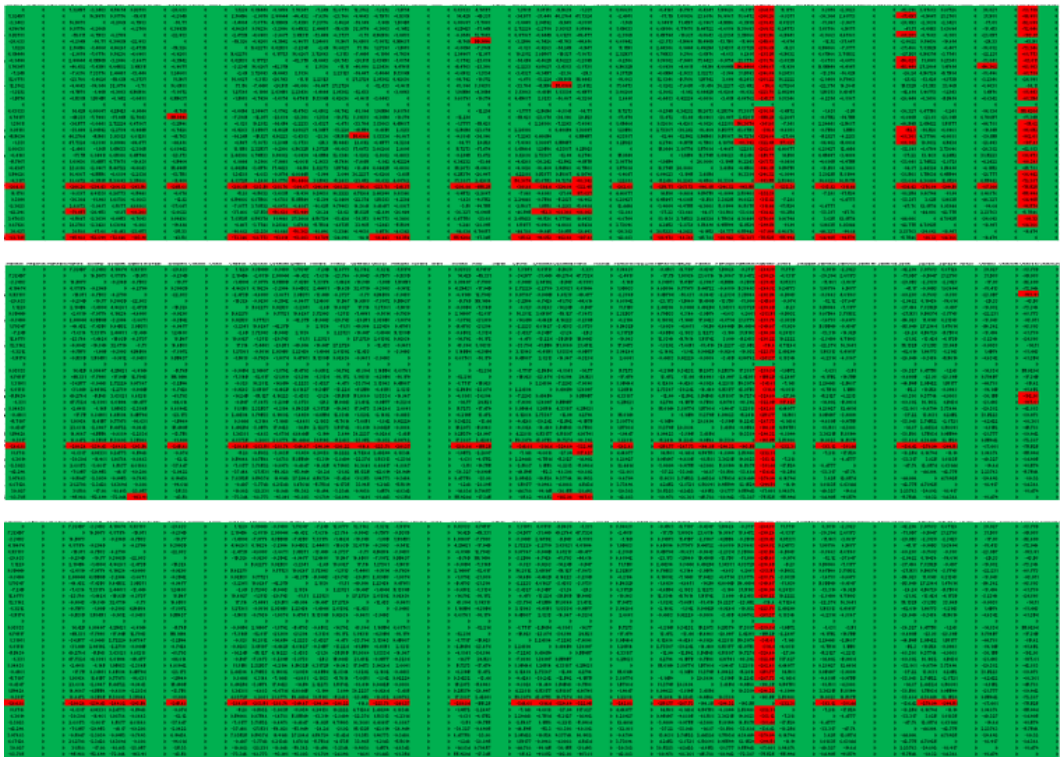
Obr. 13: Porovnanie WAVG a DoC - Rozdiel 50km, 100km, 200km

Nakoniec sme porovnali aj naše dve metódy, AVG a WAVG. Ich rozdiel by som nepovažoval za signifikantný. Vzdialenosti, ktoré ostali červené aj pri rozdieloch 200 km a 4 h, sú vzdialenosti, v ktorých je jednou z krajín Rusko. Pravdepodobnou príčinou je veľkosť tejto krajiny, obrovské vzdialenosti medzi mestami, pri ktorých aj malá zmena



Obr. 14: Porovnanie WAVG a DoC - Rozdiel 1h, 2h, 4h

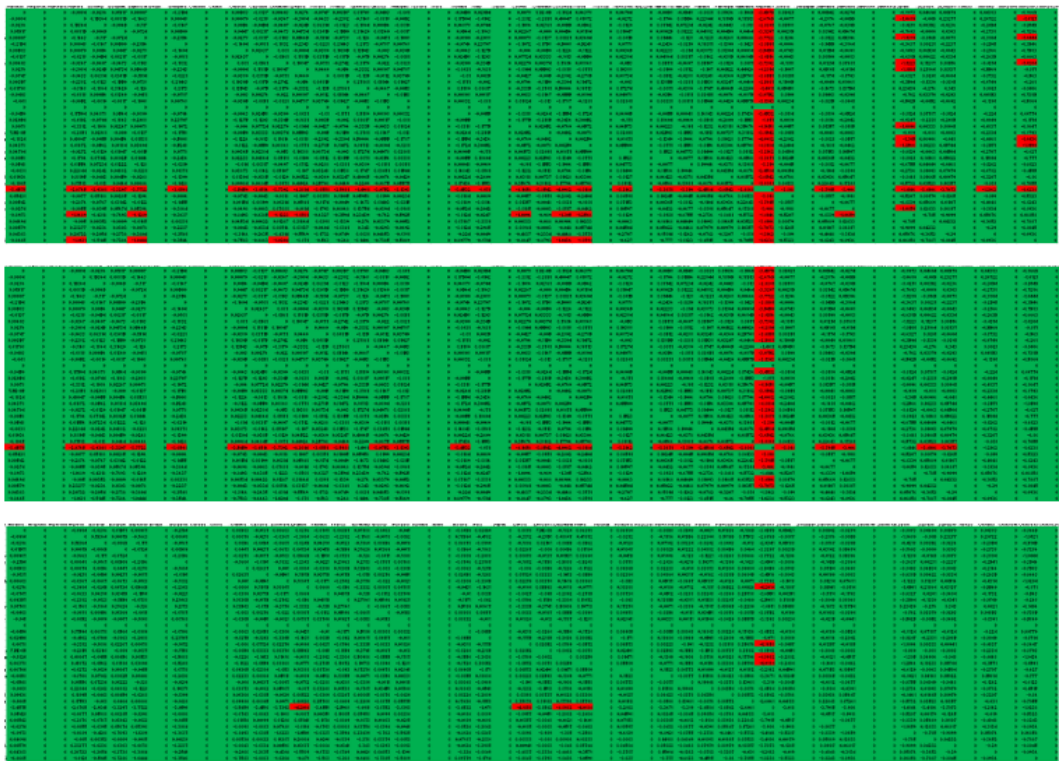
váhy vedie k signifikantnej zmene rozdielu medzi metódami. Napriek tomu možno kon-



Obr. 15: Porovnanie WAVG a AVG - Rozdiel 50km, 100km, 200km

štatovať, že rozdiely medzi týmito dvoma metódami sú malé. Porovnanie metód AVG

a WAVG ilustrujú obrázky 15 a 16.



Obr. 16: Porovnanie WAVG a AVG - Rozdiel 1h, 2h, 4h

Z dôvodu, že metódy AVG a WAVG sa veľmi nelíšia, budeme v ďalšom priebehu práce na určovanie vzdialenosti krajín využívať už iba metódu AVG. Taktiež od určitého momentu sú veľké vzdialenosti brané len ako veľké (bez ohľadu na ich vzdájomnú vzdialenosť) a rovnako v prípade malých vzdialeností krajín. Preto na AVG a AVG_čas aplikujeme logaritmickú transformáciu.

3.6 Svetový rebríček konkurencieschopnosti v digitalizácii krajiny 2018

Moderné technológie sa stále viac využívajú aj v procese výučby na vysokých školách. Pripravujú študentov lepšie na prax a otvárajú nové možnosti rozvoja. Uľahčujú administratívne úkony a fungovanie v rámci vysokej školy. Vízia futuristického štúdia láka študentov migrovať do krajín s vyššou mierou digitalizácie. Ako posledný parameter využijeme rebríček Svetovej konkurencieschopnosti digitalizácie z roku 2018. Viac informácií o tvorbe dát nájdete v zdroji [3].

4 Lineárna regresia

V nasledujúcej časti sa budeme venovať zostavovaniu prvých modelov pre podrobnejšiu analýzu dát. Napriek tomu, že sme sa pri niektorých dátach rozhodli využiť logaritmickú transformáciu pre linearizáciu dát a presnejšie zachytenie percepcie študentov, modeli sú stále považované za lineárne a preto ide o lineárnu regresiu.

V prvom kroku sme sa rozhodli rozdeliť dáta do dvoch hlavných skupín dát. Predpokladáme, že dáta z tabuľky Get Better Life by mohli mať významný vplyv na tok študentov. Tieto dáta sú zbierané pre menší počet krajín v porovnaní s ostatnými našimi dátami. Preto všetky nasledujúce modely budú riešené pre dve skupiny dát. Prvá skupina dát, zahŕňa všetky naše krajiny, bude postavená na dátach bez údajov Better Life, zatiaľ čo druhá skupina bude založená na dátach vrátane údajov Better Life, pričom krajiny tohto datasetu sú prispôsobené obsahu tejto tabuľky. Pre každý z modelov využijeme pre prvotnú analýzu tri prístupy.

Pred bližšou analýzou regresnými modelmi sme sa rozhodli normalizovať naše parametre tak, aby mali priemer rovný jednej. Zdôvodnenie normalizácie dát na priemer 1 vyplýva z potreby zabezpečiť správnu interpretáciu modelu a efektívnu konvergenciu optimalizačných algoritmov v regresnom prostredí, pričom berieme do úvahy variabilitu hodnôt v jednotlivých premenných.

Rôznorodosť hodnôt v rámci premenných môže vytvárať nevyvážený vplyv na model regresie. Normalizácia na priemer 1 zabezpečuje, že výstupy modelu sú porovnateľné, pretože každá premenná má v priemere rovnaký dosah. To poskytuje jednoduchšiu interpretáciu vplyvu každej premennej na výstup modelu, čo je kľúčové pri vysvetľovaní dát a rozhodovaní založenom na modeloch.

4.1 Modely

Nech M_{ij} označuje počet študentov odchádzajúcich z krajiny i do krajiny j, kde $i = 1, \ldots, N$ a $j = 1, \ldots, M$. N reprezentuje počet krajín v množine $ORIGINS$ pre danú dátovú sadu a M je počet krajín v množine $DESTINATION$ pre tú istú dátovú sadu. Nech ${R_{ij}}^k$ reprezentuje vzťah medzi krajinou i, kde $i = 1, \ldots, N$, a krajinou j, $j = 1, \ldots, M$, v k-tej matici vzťahov. Tieto matice vzťahov sú štyri: váženú časovú

vzdialenosť medzi krajinami, váženú geografickú vzdialenosť, minulé toky študentov a susednosť krajín. α^k označuje regresný koeficient priradený k k-tej matici vzťahov.

Nech ${Q_j}^l$ predstavuje l-tú kvalitatívnu hodnotu cieľovej krajiny j, kde $j = 1, \ldots, M$. β^l označuje regresný koeficient priradený k l-tému kvalitatívnemu parametru. Za kvalitatívny parameter považujeme akýkoľvek parameter, ktorý nebol podrobený logaritmickej transformácii.

Nech ${LQ_j}^s$ je s-tá logaritmovaná kvalitatívna hodnota cieľovej krajiny j, kde $j = 1, \ldots, M$. γ^s označuje regresný koeficient priradený k s-tému logaritmovanému kvalitatívnemu parametru. Za logaritmovaný kvalitatívny parameter považujeme akýkoľvek parameter, ktorý prešiel logaritmickou transformáciou.

Potom prvý regresný model má tvar:

$$M_{ij} = \sum_{k=1}^{K} \alpha^k {R_{ij}}^k + \sum_{l=1}^{L} \beta^l {Q_j}^l + \sum_{s=1}^{S} \gamma^s {LQ_j}^s + \epsilon \tag{6}$$

kde ϵ predstavuje náhodnú chybu modelu.

Tento prístup nám umožní zistiť priame (v tomto prípade aj lineárne) vzťahy. Z pohľadu študenta možno túto rovnicu interpretovať aj ako dôležité prvky, ktoré na danej krajine vníma pozitívne. Študenti však často neberú do úvahy len kladné stránky destinácie štúdia, ale túto krajinu porovnávajú so svojou. Aby sme zohľadnili aj tento prístup, v druhom modeli sa pozrieme na rozdiely medzi jednotlivými krajinami. Do kvalitatívnej časti pridáme kvalitatívny l-tý parameter krajiny pôvodu študenta ${Q_i}^l$, kde $i = 1, \ldots, N$. Do logaritmickej kvalitatívnej časti pridáme s-tý logaritmovaný kvalitatívny parameter krajiny pôvodu študenta ${LQ_i}^s$, kde $i = 1, \ldots, N$.

Potom druhý regresný model má tvar:

$$M_{ij} = \sum_{k=1}^{K} \alpha^k {R_{ij}}^k + \sum_{l=1}^{L} \beta^l ({Q_j}^l - {Q_i}^l) + \sum_{s=1}^{S} \gamma^s ({LQ_j}^s - {LQ_i}^s) + \epsilon \tag{7}$$

kde ϵ predstavuje náhodnú chybu modelu.

Zatiaľ sme predstavili dva regresné modely, ktoré pomáhajú analyzovať vzťahy medzi krajinami na základe rôznych kvalitatívnych a kvantitatívnych faktorov. Teraz zavedieme tretí regresný model, ktorý poskytuje iný pohľad na dynamiku medzi krajinami pôvodu a destinácie.

Predpokladajme, že interakcie medzi krajinami nie sú len lineárne alebo založené na rozdieloch, ale môžu zahŕňať aj proporčné vzťahy. Preto v tomto modeli zohľadňujeme proporcie medzi kvalitatívnymi hodnotami krajín pôvodu a destinácie. To nám umožní lepšie pochopiť, ako relatívne vlastnosti krajín ovplyvňujú vzájomný pohyb študentov.

Tretí regresný model je definovaný ako:

$$M_{ij} = \sum_{k=1}^{K} \alpha^k {R_{ij}}^k + \sum_{l=1}^{L} \beta^l \frac{{Q_j}^l}{{Q_i}^l} + \sum_{s=1}^{S} \gamma^s ({LQ_j}^s - {LQ_i}^s) + \epsilon \tag{8}$$

kde ϵ predstavuje náhodnú chybu modelu.

Tieto tri modely nám pomôžu lepšie vysvetliť klúčové faktory pri v rozhodovacom proces študentov pri výbere ich destinácie štúdia. Náš konečný model, zostavený kombináciou týchto modelov, zvýši komplexnosť celkového modelu a zohľadní rôzne motivácie študentov. Integrácia rozličných prístupov nám umožní získať ucelenejší a detailnejší pohľad na faktory, ktoré ovplyvňujú výber študijného miesta.

4.2 Vytvorenie modelu po krajinách z prvej skupiny dát

Pri analýze prvej skupiny dát zvolíme dva prístupy modelovania: globálny a lokálny. V tejto časti sa bližšie zameriame na lokálny prístup, teda budeme modelovať regresiu pre jednotlivé krajiny z množiny *ORIGINS* samostatne. Predpokladáme, že motivácia študentov odcestovať za štúdiom do zahraničia sa bude líšiť vzhľadom na krajinu z dôvodu kultúrnych, ekonomických a sociálnych rozdielov medzi krajinami. Nevýhodou tohto modelu je malý počet dát pre jednotlivé modely krajín, keďže každá regresná rovnica bude modelovaná oddelene pre krajiny z množiny *ORIGINS*, za pomoci krajín z množiny *DESTINATIONS* a k nim prislúchajúcich dát, pričom sa použije rovnica (6). Keby sme chceli určiť presné modely krajín, potrebovali by sme do vzorku pridať ďalšie krajiny a dáta k nim. Nám však v tomto modeli ide o preukázanie rozmanitosti rozhodovania študentov vzhľadom na krajinu, z ktorej pochádzajú. Takýto model bude dostatočný pre preukázanie trendov v rámci Európy, pričom pôjde prevažne o rozdiely medzi vyspelejšími a menej vyspelejšími krajinami.

Výsledky pre jednotlivé krajiny nájdete v prílohe. Pre demonštráciu rozdielnosti

regresných modelov a trendov krajín využijeme Slovensko, Švédsko, Španielsko, ako aj mapové grafy parametrov *Susedné štáty*, *TOP 200 Univerzít* a *Nezamestnanosť*. V tabuľke 1 je možné sledovať rozdielnosti jednotlivých krajín. Hranicu p hodnoty možných signifikantných prvkov sme zvolili rovnú 25%, nie 5%, z dôvodu nedostatku dát a tiež možnosti iných ako lineárnych vzťahov. Parametre, ktoré majú p hodnotu menšiu, sme označili ako potenciálne signifikantné parametre.

Zostrojená regresná rovnica z týchto parametrov by pre slovensko vyzerala nasledovne:

$$\begin{aligned}\textbf{Odlev študentov} &= \alpha_1^1\textbf{Minulé toky} + \beta_1^1\textbf{Rebríček vzdelávania} \\ &+ \beta_2^1\textbf{Výskum ako časť HDP} + \beta_3^1\textbf{Náboženská frakcionizácia} \\ &+ \gamma_1^1\textbf{Top 200 Univerzít} + \gamma_2^1\textbf{Top 200 Univerzít - body} \\ &+ \gamma_3^1\textbf{Hrubý pomer ukončenia vys. škôl} + \gamma_4^1\textbf{Populácia destinácie} + \epsilon\end{aligned} \tag{9}$$

Regresná rovnica pre Španielsko by vyzerala nasledovne:

$$\begin{aligned}\textbf{Odlev študentov} &= \alpha_0^2 + \alpha_1^2\textbf{Susedné štáty} + \beta_1^2\textbf{Priemerný príjem} \\ &+ \beta_2^2\textbf{Gini index} + \beta_3^2\textbf{Výskum ako časť HDP} + \beta_4^2\textbf{Jazyková frakcionizácia} \\ &+ \gamma_1^2\textbf{HDP v PKS} + \gamma_2^2\textbf{Nezamestnanosť} + \gamma_3^2\textbf{Top 200 Univerzít} \\ &+ \gamma_4^2\textbf{Všetky Univerzity} + \epsilon\end{aligned} \tag{10}$$

Regresná rovnica pre Švédsko by vyzerala nasledovne:

$$\begin{aligned}\textbf{Odlev študentov} &= \alpha_0^2 + \alpha_1^2\textbf{Minulé toky} + \beta_1^2\textbf{Gini index} + \beta_2^2\textbf{Rozptyl inflácie} \\ &+ \beta_3^2\textbf{Dĺžka života} + \beta_4^2\textbf{Náboženská frakcionizácia} + \gamma_1^2\textbf{HDP v PKS} \\ &+ \gamma_2^2\textbf{Top 200 Univerzít} + \gamma_3^2\textbf{Top 200 Univerzít - body} + \gamma_4^2\textbf{Všetky Univerzity} \\ &+ \gamma_5^2\textbf{Všetky Univerzity - body} + \gamma_6^2\textbf{Populácia destinácie} + \epsilon\end{aligned} \tag{11}$$

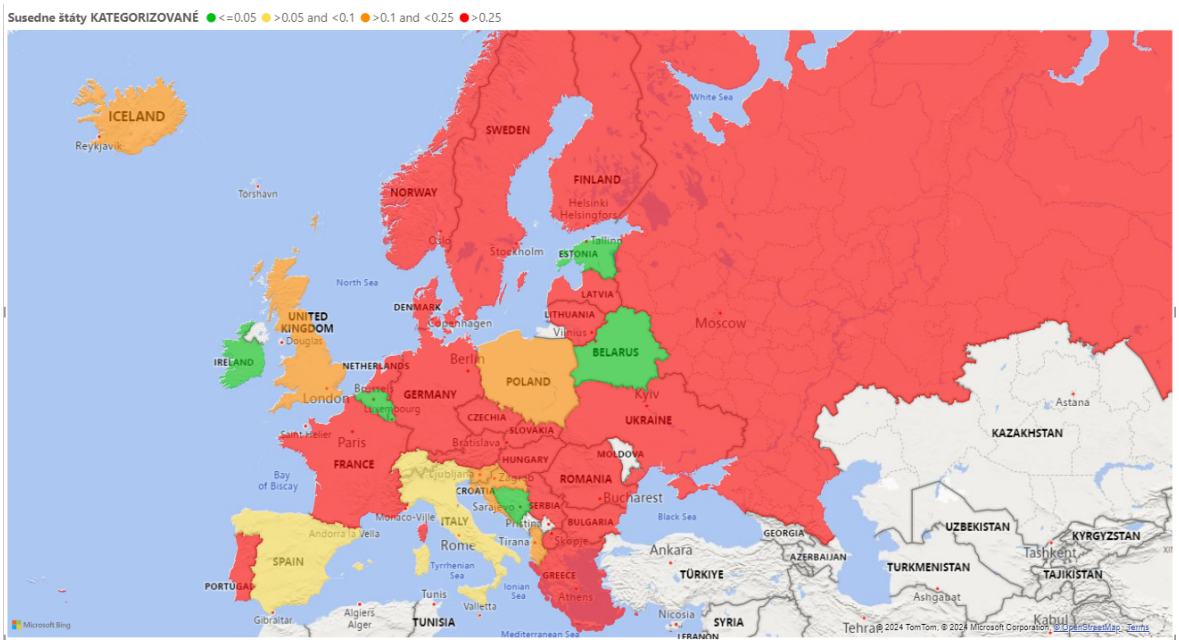
Tabuľka 1: P hodnoty po krajinách: Slovensko, Španielsko, Švédsko

Parameter	Slovakia	Spain	Sweden
Intercept	0.2409	0.346	0.1505
Minulé toky	0.0091	0.664	0.0386
Vážené vzdialenosti	0.6534	0.5105	0.6175
Vážené časové vzdialenosti	0.4831	0.3402	0.6590
Susedné štáty	0.6589	0.0691	0.5970
Rast HDP v %	0.4805	0.6844	0.7648
HDP v PKS na obyvateľa	0.5782	0.4474	0.4172
Priemerný príjem	0.7802	0.0789	0.8406
Gini index	0.9982	0.172	0.2012
Priemerná inflácia	0.2849	0.6175	0.3921
Maximálna inflácia	0.7373	0.3370	0.6651
Rozptyl inflácie	0.8120	0.299	0.1908
Rebríček vzdelávania	0.0999	0.4691	0.8372
Výskum ako časť HDP	0.2298	0.0705	0.5927
Politická stabilita	0.6653	0.3434	0.8591
Dĺžka života	0.3124	0.579	0.0864
Rebríček korupcie	0.6072	0.8120	0.4589
Etická frakcionizácia	0.8617	0.8756	0.3768
Jazyková frakcionizácia	0.3385	0.1054	0.7777
Náboženská frakcionizácia	0.1075	0.905	0.0926
HDP v PKS	0.7967	0.105	0.1861
Nezamestnanosť	0.7269	0.0662	0.8607
Výdavky štátu na vys. školy % HDP	0.9192	0.5518	0.2573
TOP 200 Univerzít	0.1760	0.000	0.0025
TOP 200 Univerzít - body	0.0600	0.657	0.2476
Všetky Univerzity	0.6338	0.027	0.1648
Všetky Univerzity - body	0.3437	0.099	0.0766
Hrubý pomer ukončenia vys. škôl	0.2020	0.8806	0.6402
Populácia destinácie	0.2409	0.346	0.1505

Z tabuľky 1 a z porovnania rovníc (9), (10) a (11) vyplýva rozdielnosť modelov pre jednotlivé krajiny. Zaujímavé je sledovať opakovanie niektorých parametrov, čo by mohlo naznačovať existenciu účinného globálneho modelu pre odlev mozgov medzi krajinami. Taktiež to poukazuje na existenciu trendov v rámci európskych krajín.

Pre bližšiu analýzu sme vybrané parametre vizualizova v prostredí Power BI. Obrázok 17 zobrazuje p hodnotu parametra Susedné štáty pre jednotlivé štáty Európy. Tieto hodnoty sme rozdelili do troch skupín:

- P hodnoty menšie rovné ako 0,05 sú označené zelenou farbou
- P hodnoty z intervalu (0,05; 0,1] sú označené žltou farbou
- P hodnoty z intervalu (0,1; 0,25] sú označené oranžovou farbou
- P hodnoty väčšie ako 0,25 sú označené červenou farbou

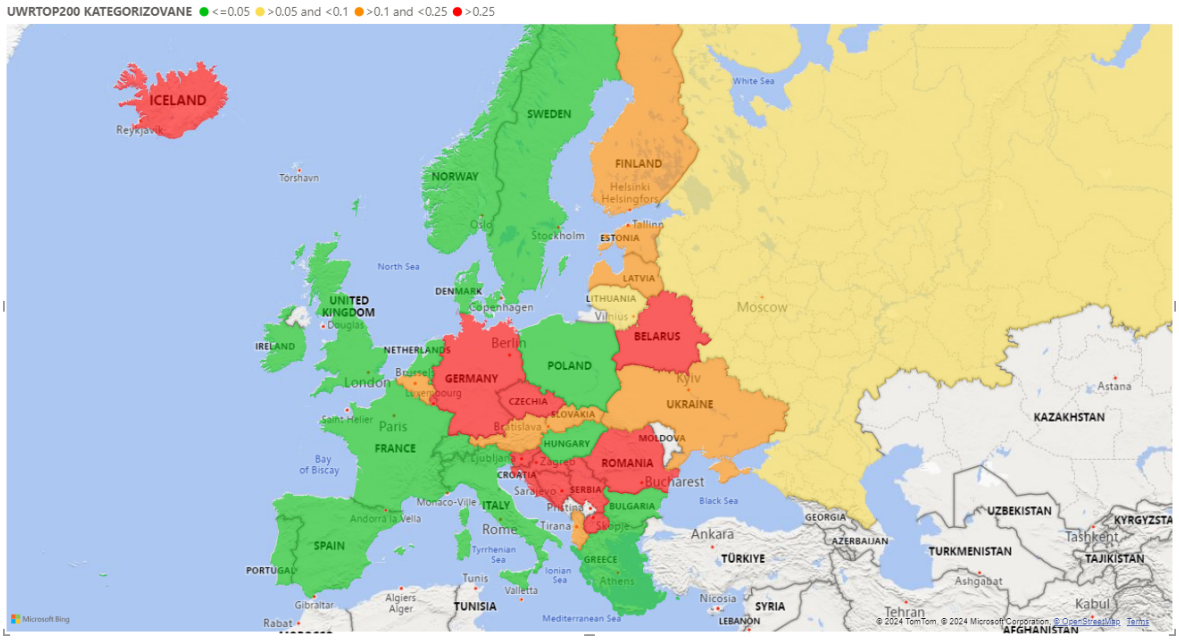


Obr. 17: P-value parametra Susedné štáty v modeli po krajinách

Parameter susedných krajín na prvý pohľad nevykazuje výrazné trendy v odleve mozgov. Avšak, pri podrobnejšej analýze tejto premennej možno pozorovať niekoľko zaujímavých skutočností. Zistili sme, že krajiny, pre ktoré parameter susedných štátov dosahuje hodnoty p menšie ako 25%, sú zväčša tie, ktoré majú malý počet susedných

krajín. Tento jav si zasluhuje ďalšie preskúmanie, aby sa lepšie pochopili mechanizmy, ktoré stoja za týmito trendmi v odleve mozgov v kontexte globálnejšej perspektívy.

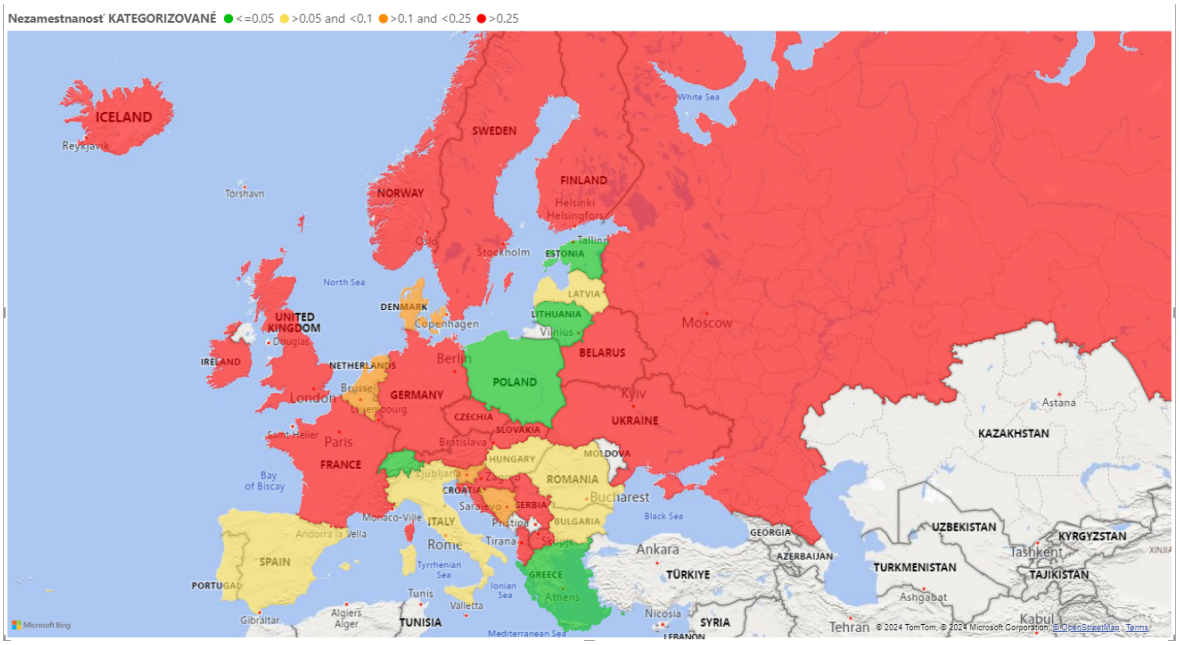
Druhým parametrom, ktorý sme si vybrali na podrobnejšie skúmanie, je TOP 200 univerzít. Už na prvý pohľad je zrejmé, že tento parameter predstavuje významný ukazovateľ pre väčšinu krajín. Pri dôkladnejšej analýze je možné vypozorovať jeho význam najmä pre vyspelé krajiny západnej Európy a škandinávske krajiny. Prekvapením sú Poľsko a Maďarsko, kde pre študentov zohráva štúdium v krajinách s prestížnymi školami tiež dôležitú úlohu. V prípade Poľska je zreteľná najmä migrácia do Veľkej Británie a Spojených štátov, ktorá je úzko spojená aj s pracovnou migráciou. V Maďarsku je naopak zvýšený záujem študentov o štúdium nielen vo Veľkej Británii, ale aj v Nemecku, Holandsku a Rakúsku. Tieto zistenia naznačujú, že prítomnosť univerzít z TOP 200 má značný vplyv na vzory vzdelávacej a zároveň pracovnej migrácie v týchto regiónoch.



Obr. 18: P-value parametra TOP 200 Univerzít v modeli po krajinách

Vzhľadom na vysokú signifikantnosť tohto parametra pre viaceré európske krajiny predpokladáme, že by mohol mať signifikantný vplyv aj v globálnom modeli odlevu študentov v neskorších fázach našej práce. Táto hypotéza podčiarkuje potenciálnu globálnu relevanciu prítomnosti prestížnych vysokých škôl ako faktora ovplyvňujúceho migráciu študentov.

Posledným parametrom, ktorý sme sa rozhodli podrobne skúmať, je ekonomický ukazovateľ nezamestnanosti. Zistili sme, že koncentrácia krajín, pre ktoré je tento parameter signifikantný, je vyššia vo východnej Európe. Príčinou je nižšia ekonomická vyspelosť týchto štátov, čo v percepcii študentov odráža motiváciu hľadať lepšie príležitosti v ekonomicky rozvinutejších krajinách.



Obr. 19: P-value parametra Nezamestnanosť v modeli po krajinách

Signifikantnosť nezamestnanosti je tiež pozorovateľná v krajinách, ktoré v minulosti čelili ekonomickej kríze alebo boli jej blízko, ako sú napríklad Španielsko, Taliansko a Grécko. Tento zistený vzťah poukazuje na to, ako ekonomické faktory, a konkrétne nezamestnanosť, môžu ovplyvňovať migračné rozhodnutia študentov a naznačuje dôležitosť ekonomických podmienok v krajinách pôvodu pri štúdiu globálnych migračných trendov.

4.3 Vytvorenie globálneho modelu prvej skupiny dát

Lokálny model odhalil rozdiely v rozhodovacích procesoch študentov medzi jednotlivými krajinami. Avšak kvôli nedostatku dát tento model nie je vhodný pre podrobnú analýzu. Súčasne však poukázal na existenciu spoločných trendov medzi krajinami zoskupenými v súbore ORIGINS. Naším cieľom je vytvoriť globálny model obsahujúci signifikantné parametre, ktoré sú kľúčové pre celú Európu. Prednosťou tohto mo-

delu je možnosť odôvodniť migračné rozhodnutie náhodne vybraného študenta. Vďaka neskoršej aplikácii metódy rozhodovacích stromov zachováme rozmanitosť rozhodovania študentov z rôznych krajín.

Uvedomujeme si, že percepcia študentov vzhľadom na rôzne parametre sa líši a predpokladáme aj existenciu ne-lineárnych vplyvov. Preto na analýzu počiatočných dát aplikujeme tri rôzne prístupz popísané rovnicami (6), (7), (8), čím získame tri rozdielne pohľady na analýzu dát.

Ako prvé aplikujeme na analýzu rovnicu (6), pričom výsledky sú uvedené v tabuľke 2. Keďže naša práca na identifikácii najlepšieho modelu s využitím rovnice (6) ešte nie je finálna, do ďalšej fázy zahrnieme všetky parametre s p-hodnotou nižšou ako 25%. Výnimky nastanú, ak z jednej kategórie vyplynie viacero parametrov s p-hodnotou pod 25%, ako napríklad u „Vážených vzdialeností“. V takýchto prípadoch do ďalšej analýzy zahrnieme iba parameter s nižšou p-hodnotou. Konkrétne sme vybrali parametre: Minulé toky študentov, Vážená časová vzdialenosť, Susedné štáty, Priemerná inflácia, Výskum ako percento HDP, Jazyková fragmentácia, HDP na obyvateľa, Nezamestnanosť, TOP 200 univerzít, Všetky univerzity, Všetky univerzity - body a Hrubý pomer ukončenia vysokoškolského štúdia.

Pri ďalšom skúmaní vybraných parametrov sme sa snažili identifikovať najlepší submodel. Tento proces zahŕňal analýzu, či majú niektoré z lineárnych parametrov v skutočnosti nelineárny vplyv pre odlev mozgov. Na takéto parametre aplikujeme logaritmickú transformáciu. Medzi vybranými parametrami sú tri lineárne parametre: priemerná inflácia, výskum ako percento HDP a jazyková fragmentácia. Pri tvorbe submodelov sme postupne logaritmovali každý z týchto parametrov a tiež experimentovali s kombináciami logaritmovaných parametrov. Za najlepší model sme považovali ten, ktorý obsahoval najvyšší počet signifikantných parametrov.

Tabuľka 2: Prístup 1 - vstupná regresia

Parameter	Odhad	Štand. odchýlka	P - value
(Intercept)	5106.944	3350.423	1.276×10^{-1}
Minulé toky študentov	1396.678	92.642	$<2.000\times10^{-16}$
Vážená vzdialenosť	−467.099	378.284	2.171×10^{-1}
Vážená časová vzdialesnoť	621.708	347.564	7.383×10^{-2}
Susedné krajiny	1340.222	181.794	2.630×10^{-13}
Rast HDP v %	6.812	117.720	9.539×10^{-1}
HDP v PKS na obyvateľa	−186.940	275.527	4.976×10^{-1}
Priemerný príjem	237.749	213.441	2.655×10^{-1}
Gini index	319.814	303.519	2.922×10^{-1}
Priemerná inflácia	340.904	173.842	5.005×10^{-2}
Maximálna inflácia	−292.120	238.435	2.207×10^{-1}
Rozptyl inflácie	45.634	68.813	5.073×10^{-1}
Rebríček vzdelávania	−1172.045	1200.210	3.289×10^{-1}
Výskum ako časť HDP	−220.124	126.760	8.265×10^{-2}
Politická stabilita	−33.617	86.339	6.971×10^{-1}
Dĺžka života	−694.846	2491.387	7.804×10^{-1}
Rebríček korupcie	−77.532	170.882	6.501×10^{-1}
Etická frakcionizácia	150.770	117.445	1.994×10^{-1}
Jazyková frakcionizácia	−175.959	104.835	9.345×10^{-2}
Náboženská frakcionizácia	−40.252	168.181	8.109×10^{-1}
HDP v PKS	−2627.250	1402.807	6.126×10^{-2}
Nezamestnanosť	−1071.928	275.692	1.050×10^{-4}
Výdavky štátu na vys. školy % HDP	−24.998	1417.800	9.859×10^{-1}
TOP 200 Univerzít	301.916	58.586	2.860×10^{-7}
TOP 200 Univerzít - body	58.846	118.946	6.209×10^{-1}
Všetky Univerzity	1458.729	531.650	6.138×10^{-3}
Všetky Univerzity body	−1461.101	589.099	1.323×10^{-2}
Hrubý pomer ukončenia vys. škôl	−1096.807	782.321	1.611×10^{-1}

Počiatočný model, ktorý pozostáva iba z vybraných parametrov a nezahŕňa logaritmickú transformáciu, obsahoval desať signifikantných parametrov a jeden parameter, ktorého p-hodnota bola blízko hranice signifikantnosti, teda hladiny významnosti 5 %. V tabuľke 3 nájdete priebeh testovania submodelov. Stĺpec „***“ označuje počet parametrov s p-hodnotou menšou ako 0.001, stĺpec „**“ s p-hodnotou menšou ako 0.01, stĺpec „**“ s p-hodnotou menšou ako 0.05 a stĺpec „.“ s p-hodnotou menšou ako 0.1. Následne sme sledovali modely, pri ktorých došlo k zlepšeniu. Zlepšenie možno pozorovať pri modeloch Submodel02, Submodel03 a Submodel05, pričom pri modeli Submodel02 je počet signifikantných parametrov 11 a pri Submodel03 a Submodel05 je počet signifikantných parametrov 12. Preto najlepší model budeme vyberať spomedzi Submodelu03 a Submodel05.

Tabuľka 3: Hľadananie najlepšieho submodelu k prvému prístupu - logaritmácia lineárnych parametrov

Submodel	***	**	*	.	**Poznámka**
Submodel01	6	2	2	1	Model s logaritmizovanou priemernou infláciou
Submodel02	6	4	1	1	Model s logaritmizovaným výskumom ako percentom HDP
Submodel03	8	3	1	0	Model s logaritmizovanou jazykovou fragmentáciou
Submodel04	5	0	4	2	Model s logaritmizovanou priemernou infláciou a výskumom ako percentom HDP
Submodel05	5	5	2	0	Model s logaritmizovaným výskumom ako percentom HDP a jazykovou fragmentáciou
Submodel06	6	2	2	2	Model s logaritmizovanou priemernou infláciou a jazykovou fragmentáciou
Submodel07	5	0	5	2	Model s logaritmizovanou priemernou infláciou, výskumom ako percentom HDP a jazykovou fragmentáciou
Submodel08	8	3	1	0	Model s logaritmizovanou jazykovou fragmentáciou, bez započítania váženej časovej vzdialenosti

V záverečnej fáze sme z týchto modelov vylúčili nesignifikantné parametre. V oboch modeloch sa ako jediný nesignifikantný parameter ukázala „Vážená časová vzdialenosť“. Po jej odstránení sme za lepší vybrali Submodel03, uvedený v tabuľke 4, ktorý obsahuje väčší počet veľmi signifikantných parametrov. Tento model budeme používať v ďalšom priebehu hľadania najlepšieho globálneho modelu prvej skupiny dát.

Tabuľka 4: Submodel 3 - najlepší zo submodelov

Parameter	Odhad	Štandardná odchýlka	P - value
Intercept	2706.39	767.23	4.310×10^{-4}
Minulé toky študentov	1384.23	87.73	$<2.000\times10^{-16}$
Vážená časová vzdialesnoť	82.81	67.41	2.195×10^{-1}
Susedné krajiny	1192.39	159.95	1.430×10^{-13}
Priemerná inflácia	124.98	41.96	2.937×10^{-3}
Výskum ako časť HDP	−370.99	101.43	2.620×10^{-4}
Logaritmovaná jazyková frakcionizácia	92.32	41.70	2.699×10^{-2}
HDP v PKS	−1648.67	584.72	4.865×10^{-3}
Nezamestnanosť	−742.74	167.55	9.890×10^{-6}
TOP 200 Univerzít	276.07	39.21	2.770×10^{-12}
Všetky Univerzity	1219.91	345.32	4.220×10^{-4}
Všetky Univerzity - body	−1450.39	375.42	1.160×10^{-4}
Hrubý pomer ukončenia vys. škôl	−1255.44	418.02	2.710×10^{-3}

Tabuľka 5: Submodel08 - finálny model z 1. prístupu, konštruovaný zo Submodelu03 po odstránení Váženej časovej vzdialenosti

Parameter	Odhad	Štandardná odchýlka	P - value
Intercept	2565.16	758.68	7.380×10^{-4}
Minulé toky študentov	1407.76	85.62	$<2.000\times10^{-16}$
Susedné krajiny	1167.98	158.74	2.900×10^{-13}
Priemerná inflácia	117.52	41.52	4.708×10^{-3}
Výskum ako časť HDP	−390.81	100.16	9.910×10^{-5}
Logaritmovaná jazyková frakcionizácia	99.34	41.32	1.631×10^{-2}
HDP v PKS	−1560.87	580.43	7.233×10^{-3}
Nezamestnanosť	−694.22	162.85	2.130×10^{-5}
TOP 200 Univerzít	273.20	39.15	4.250×10^{-12}
Všetky Univerzity	1109.53	333.48	8.960×10^{-4}
Všetky Univerzity - body	−1385.68	371.77	2.000×10^{-4}
Hrubý pomer ukončenia vys. škôl	−1118.95	403.04	5.559×10^{-3}

Druhý prístup spočíva v použití rovnice (7). Najprv sme upravili dátový súbor tak, aby jednotlivé riadky regresného modelu obsahovali rozdiely parametrov, pričom rozdiely sme počítali len u kvalitatívnych parametrov rovnice. Následne je postup identický s postupom použitým v prvej časti. Najprv vyberieme potenciálne signifikantné parametre s p hodnotou menšou ako 0.25, a potom budeme vytvárať submodely, pričom na lineárne dáta aplikujeme logaritmickú transformáciu.

Tabuľka 6: Hľadananie najlepšieho submodelu k druhému prístupu

Submodel	***	**	*	.	**Poznámka**
Submodel11	4	3	1	1	Počiatočný submodel
Submodel12	5	2	1	1	Model s logaritmizovaným rozdielom v etickej fragmentácii
Submodel13	4	3	1	1	Model s logaritmizovaným rozdielom v dĺžke života
Submodel14	4	2	2	1	Model s logaritmizovanoým rozdielom skóre korupcie
Submodel15	5	3	2	0	Model s logaritmizovaným rozdielom Maximálnej inflácie
Submodel16	5	3	1	0	Najlepší model po vynechaní nesignifikantných parametrov

Pri konštrukcii submodelov v tejto časti nedošlo k výraznému zlepšeniu modelu po logaritmovaní jednotlivých parametrov, ako je vidieť v tabuľke 6. Zdá sa, že najlepší model, je Submodel15, tabuľka 7. Model ale spôsobil zhoršenie p-hodnoty práve pre logaritmovaný parameter maximálnej inflácie, pre ktorý bol tento submodel vytvorený. Z tohto dôvodu nebudeme brať Submodel15 ako najlepší zo submodelov.

Tabuľka 7: Submodel 15 - poukázanie na fakt že p hodnota logaritmovaného parametra sa výrazne zhoršil oproti počiatočnému Submodelu 01

Parameter	**Odhad**	**Štandardná odchýlka**	**P - value**
(Intercept)	−1046.66	108.05	$<2.000\times10^{-16}$
Minulé toky	1417.87	72.25	$<2.000\times10^{-16}$
Vážený časový odstup	41.48	65.38	5.259×10^{-1}
Susedné krajiny	1227.17	157.20	1.040×10^{-14}
Log. Rozdiel maximálnej inflácie	0.97	78.98	9.902×10^{-1}
Rozdiel v dĺžke života	−2396.84	945.56	1.134×10^{-2}
Rozdiel v hodnotenie korupcie	133.28	107.85	2.167×10^{-1}
Rozdiel v etickej frakcionizácii	68.54	61.26	2.634×10^{-1}
Rozdiel HDP v PKS	−1395.63	673.56	3.842×10^{-2}
Rozdiel nezamestnanosť	−427.07	123.93	5.800×10^{-4}
Rozdiel vo výdavkoch štátu na vys. školy % HDP	581.75	809.47	4.724×10^{-1}
Rozdiel v TOP 200 Univerzitách	84.36	29.22	3.930×10^{-3}
Rozdiel vo všetkých Univerzitách	625.21	212.40	3.290×10^{-3}
Rozdiel vo všetkých Univerzitách - body	−685.50	256.04	7.500×10^{-3}

Tabuľka 8: Submodel11 - počiatočný model

Parameter	**Odhad**	**Štandardná odchýlka**	**P - value**
Intercept	−1039.75	106.72	$<2.000\times10^{-16}$
PastFlows_loop	1410.98	72.19	$<2.000\times10^{-16}$
WeightedTimeDistance_model1_loop	41.29	64.90	5.247×10^{-1}
neighbouring_countries_model1_loop	1232.91	157.20	7.860×10^{-15}
Inflation_MAX	−25.71	21.95	2.417×10^{-1}
Life_expectancy	−2804.80	857.46	1.094×10^{-3}
CoruptionRanking	138.37	107.87	1.998×10^{-1}
Ethic	56.12	60.59	3.545×10^{-1}
GDP_PPP	−1303.03	677.87	5.475×10^{-2}
unemployment	−431.71	123.88	5.050×10^{-4}
GetePPPmil	668.44	808.91	4.087×10^{-1}
UWRTOP200	81.07	28.79	4.927×10^{-3}
UWRTOPALL	540.40	220.28	1.426×10^{-2}
UWRTOPALLpoints	−666.87	255.25	9.068×10^{-3}

Preto sme sa v prípade druhého prístupu rozhodli ako najlepší vybrať počiatočný model, tabuľka 8 a postupne odstraňovať nesignifikantné parametre, začínajúc od najviac nesignifikantného v poradí: vážená časová vzdialenosť, rozdiel vo výdavky vlády na vys. školy (PKS), rozdiel v etickej frakcionalizácii, rozdiel v rebríčku korupcie a rozdiel v maximálna iflácia. Tento postup sa ukázal byť úspešný, keďže počet signifikantných parametrov zostal nezmenený a štandardná odchýlka parametrov sa znížila, tabuľka 9.

Tabuľka 9: Submodel 16 - finálny model 2. prístupu, konštruovaný z počiatočného submodelu a odstránením nesignifikantných parametrov

Parameter	**Odhad**	**Štandardná odchýlka**	**P - value**
(Intercept)	−993.28	81.24	$<2.000\times10^{-16}$
Minulé toky	1417.27	71.75	$<2.000\times10^{-16}$
Susedné krajiny	1218.68	156.67	1.280×10^{-14}
Rozdiel v dĺžke života	−2921.80	691.45	2.510×10^{-5}
Rozdiel HDP v PKS	−1006.41	345.62	3.640×10^{-3}
Rozdiel nezamestnanosť	−359.29	112.70	1.460×10^{-3}
Rozdiel v TOP 200 Univerzitách	98.71	26.64	2.200×10^{-4}
Rozdiel vo všetkých Univerzitách	567.35	195.45	3.750×10^{-3}
Rozdiel vo všetkých Univerzitách - body	−592.12	240.51	1.392×10^{-2}

Prístup tri, za pomoci rovnice (8), nepriniesol nové parametre, ktoré by sme mohli pridať do globálneho modelu. Preto tento prístup nevyužijeme pre ďalšie hľadanie globálneho modelu.

V poslednom kroku zostrojíme globálny model prvej skupiny dát. Globálny model bude kombinovať dva prístupy, ktoré sme si vybrali pre podrobnejšiu analýzu vplyvov parametrov. Ak sa parameter ukáže ako signifikantný v oboch prístupoch, zvolíme jeho formu z prvého prístupu.

Pre vytvorenie finálneho globálneho modelu potrebujeme odstrániť nesignifikantné parametre s p-hodnotou väčšou ako $0,5$. Konkrétne odstránime parametre: intercept, rozdiel HDP v parite kúpnej sily, hrubý pomer ukončenia vysokých škôl a kvôli vysokej korelácii medzi parametrami Všetky univerzity a Všetky univerzity - body rovnú $0,916$, odstránime aj parameter Všetky univerzity.

Finálna verzia globálneho modelu prvej skupiny dát bude mať tvar:

Tabuľka 10: Globálny model prvej skupiny dát - spolu s nesignifikantnými parametrami

Parameter	Odhad	Štandardná odchýlka	P - value
(Intercept)	909.63	496.74	6.725×10^{-2}
Minulé toky	1312.23	96.03	$<2.000\times10^{-16}$
Susedné krajiny	1277.87	162.90	7.760×10^{-15}
Rozdiel HDP v PKS	−334.84	177.74	5.976×10^{-2}
Rozdiel v dĺžke života	−1385.88	622.65	2.617×10^{-2}
Výskum ako časť HDP	−428.72	100.15	1.970×10^{-5}
Log. jazyková frakcionizácia	112.99	41.97	7.170×10^{-3}
Nezamestnanosť	−620.91	159.37	1.000×10^{-4}
TOP 200 Univerzít	264.86	38.25	6.270×10^{-12}
Všetky Univerzity	483.41	208.89	2.078×10^{-2}
Všetky Univerzity - body	−1090.61	367.81	3.070×10^{-3}
Hrubý pomer ukončenia vys. škôl	−504.12	363.37	1.655×10^{-1}

Tabuľka 11: Globálny model prvej skupiny dát - finálny

Parameter	Odhad	Štandardná odchýlka	P - value
Minulé toky	1385.93	85.91	<2e-16
Susedné krajiny	1234.88	160.80	2.720×10^{-14}
Rozdiel dĺžky života	−1651.17	612.95	7.136×10^{-3}
Výskum ako časť HDP	−446.29	92.85	1.680×10^{-6}
Log. jazyková frakcionizácia	120.60	41.80	3.964×10^{-3}
Nezamestnanosť	−413.83	95.78	1.650×10^{-5}
TOP 200 Univerzít	245.87	34.12	8.790×10^{-13}
Všetky Univerzity - body	−480.01	141.55	7.120×10^{-4}

$$\begin{aligned}\textbf{Odlev študentov} = {} & \alpha_0^2 + \alpha_1^2\textbf{Minulé toky} + \alpha_2^2\textbf{Susedné krajiny} + \\ & + \beta_1^2\textbf{Rozdiel dĺžky života} + \beta_2^2\textbf{Výskum ako časť HDP} + \\ & + \beta_3^2\textbf{Log. jazyková frakcionalizácia} + \gamma_1^2\textbf{Nezamestnanosť} + \gamma_2^2\textbf{TOP 200 Univerzít} + \\ & + \gamma_3^2\textbf{Všetky Univerzity - body} + \epsilon \end{aligned} \tag{12}$$

V ďalšom priebehu práce práve na model (12) aplikujeme metódu rozhodovacích stromov, aby sme opísali rozhodovací postup študenta danej krajiny.

4.3.1 Vysvetlenie parametra Minulé toky študentov

Minulé toky študentov sa ukázali ako významný parameter pre globálny model prvej skupiny dát týkajúci sa odlivu študentov. Vysvetlenie významu týchto tokov však nie je priamočiare. V tejto časti sa pomocou lineárneho regresného modelu pokúsime odhaliť, čo všetko tento parameter obsahuje. Hoci sme váženým vzdialenostiam – časovým aj geografickým – venovali celú kapitolu 3.5 v našej práci, neukázali sa ako dôležitý determinant v globálnom modeli úniku mozgov do zahraničia, napriek ich zastúpeniu aj v odpovediach dotazníka pre stredoškolských študentov 2. Predpokladáme, že tento parameter bude zahrnutý v parametri minulých tokov.

Pomocou programovacieho jazyka R zostrojíme lineárny regresný model, ktorý bude vychádzať z rovnice (6). Oproti globálnemu modelu prvej skupiny dát sa tento model bude líšiť tým, že bude zameraný na minulé toky študentov, nie na ich odlivy medzi krajinami. V prvom kroku sme z modelu odstránili signifikantné parametre pôvodného globálneho modelu, vrátane susedných krajín, rozdielov v dĺžke života, podielu výskumu na HDP, jazykovej frakcionizácie, nezamestnanosti, parametra TOP 200 univerzít a celkového bodového hodnotenia všetkých univerzít. Regresný model minulých tokov študentov zostrojíme nasledovne:

$$\begin{aligned}
&\textbf{Minulé toky študentov} = \alpha_0^2 + \alpha_1^2\textbf{Vážená geografická vzdialenosť} + \\
&+ \alpha_2^2\textbf{Vážená časová vzdialenosť} + \beta_1^2\textbf{Rast HDP} + \beta_2^2\textbf{HDP na obyvateľa} + \\
&+ \beta_3^2\textbf{Priemerný príjem} + \beta_4^2\textbf{Gini index} + \beta_5^2\textbf{Priemerná inflácia} + \\
&+ \beta_6^2\textbf{Maximálna inflácia} + \beta_7^2\textbf{Rozptyl inflácie} + \beta_8^2\textbf{Rebríček vzdelávania} + \\
&+ \beta_9^2\textbf{Politická stabilita} + \beta_{10}^2\textbf{Rebríček korupcie} + \beta_{11}^2\textbf{Etická frakcionizácia} + \\
&+ \beta_{12}^2\textbf{Náboženská frakcionizácia} + \gamma_1^2\textbf{HDP v PKS} + \\
&+ \gamma_2^2\textbf{Výdavky vlády na vys. školy} + \gamma_3^2\textbf{Hrubý pomer ukonč. vys. škôl} + \epsilon
\end{aligned} \tag{13}$$

Výsledky regresného modela sú:

Tabuľka 12: Výsledky regresného modelu Minulých tokov - zelenou sú zvýraznené riadky, ktoré po očistení modelu od nesignifikantných prvkov ostali signifikantné

Parameter	**Odhad**	**Štandardná odchýlka**	**P - value**
(Intercept)	−2.354 29	0.213 52	<2e-16
Vážená geografická vzdialenosť	1.300 14	0.089 81	<2e-16
Vážená časová	−1.049 00	0.083 86	<2e-16
Rast HDP	−0.027 51	0.025 43	2.795×10^{-1}
HDP na obyvateľa (PKS)	0.113 92	0.060 55	6.008×10^{-2}
Priemrný príjem (PKS)	0.108 15	0.057 83	6.163×10^{-2}
Gini index	0.101 94	0.068 17	1.350×10^{-1}
Priemerná inflácia	−0.163 05	0.039 85	4.500×10^{-5}
Maximálna inflácia	0.073 44	0.059 21	2.150×10^{-1}
Rozptyl inflácie	−0.003 20	0.015 64	8.379×10^{-1}
Rebríček vzdelávania	0.993 78	0.247 98	6.400×10^{-5}
Politická stabilita	−0.021 38	0.018 76	2.546×10^{-1}
Rebríček korupcie	−0.035 58	0.047 90	4.578×10^{-1}
Etická frakcionizácia	−0.014 63	0.024 66	5.533×10^{-1}
Náboženská frakcionizácia	−0.096 08	0.033 42	4.090×10^{-3}
HDP v PKS	−0.780 91	0.252 79	2.040×10^{-3}
Výdavky vlády na vys. školy (PKS)	2.619 12	0.335 90	1.100×10^{-14}
Hrubý pomer ukonč. vys. škôl	0.247 16	0.188 32	1.895×10^{-1}

Postupným odstránením nesignifikantných parametrov dostávame finálne výsledky parametrov obsiahnutých v minulých tokoch študentov.

Tabuľka 13: Finálna tabuľka signifikantných parametrov minulých tokov

Parameter	Odhad	Štandardná odchýlka	P - value
(Intercept)	−2.341 21	0.144 63	<2e-16
Vážená geografická vzdialenosť	1.257 92	0.085 73	<2e-16
Vážená časová vzdialenosť	−1.014 82	0.081 44	<2e-16
HDP na obyvateľa (PKS)	0.151 05	0.032 16	2.870×10^{-6}
Priemerná inflácia	−0.073 86	0.010 39	1.730×10^{-12}
Rebríček vzdelávania	1.176 74	0.161 68	5.150×10^{-13}
Náboženská frakcionizácia	−0.065 69	0.024 68	7.847×10^{-3}
HDP v PKS	−0.860 89	0.231 67	2.090×10^{-4}
Výdavky vlády na vys. školy (PKS)	2.782 96	0.300 91	<2e-16

Tieto výsledky ilustrujú rozsah parametrov zahrnutých v minulých tokoch študentov. Okrem spomínaných vážených vzdialeností zahŕňajú minulé toky aj parametre ako HDP na obyvateľa v parite kúpnej sily, priemernú infláciu krajiny v rokoch 2009 - 2019, rebríček vzdelávania, náboženskú frakcionalizáciu, HDP krajiny v parite kúpnej sily a výdavky vlády na vysoké školy v parite kúpnej sily.

Minulé toky študentov predstavujú kombináciu určitých ekonomických, vzdelávacích, sociálnych, a geografických parametrov, ktoré boli dôvodmi migrácie študentov v minulosti. Možno ich chápať ako trend migrácie študentov, pričom ich signifikantnosť v globálnom modeli naznačuje, že mnohé z týchto trendov stále pretrvávajú. Z pohľadu riešenia fenoménu odlivu mozgov študentov nám tieto parametre ukazujú, že účinky aplikovaných opatrení na zníženie odlivu mozgov do zahraničia sa prejavia až po určitom čase, a to minimálne po období trvajúcom 8 rokov, zložených s nami skúmaného obdobia (štyri roky 2016 - 2019) a obdobia minulých tokov (štyri roky 2012 - 2015).

Zníženie migrácie študentov a zmena trendu odlivu mozgov nie je možné dosiahnuť na krátkodobom horizonte, ale vyžaduje si postupnú a plánovanú prácu na zlepšovaní krajiny v oblastiach, ktoré vedú študentov k migrácii za štúdiom do zahraničia. Je potrebná spolupráca viacerých vládnych období a dobrá stratégia. Je však dôležité si uvedomiť, že tieto parametre boli identifikované na základe globálneho modelu odlivu mozgov, a preto sa zloženie minulých tokov môže v jednotlivých krajinách

mierne líšiť, vzhľadom na rôznorodosť rozhodovacích procesov študentov s rozličnými štátnymi príslušnosťami.

4.3.2 Vytvorenie submodelu s jazykovými vzdialenosťami

V našej analýze nám chýbal parameter, ktorý by bližšie popisoval jazykové vzdialenosti medzi dvoma krajinami. Napríklad študenti migrujúci zo Slovenska do veľkej miery smerujú do Českej republiky z dôvodu podobnosti jazyka a schopnosti dorozumieť sa v tomto jazyku. Podobných prípadov je vo svete viacero. Parameter jazykových vzdialeností sa nám podarilo vytvoriť pre 26 krajín z našej vzorky. Aby sme nemuseli vytvárať vlastné modely špeciálne pre tento parameter, rozhodli sme sa vytvoriť submodel z Globálneho modelu prvej skupiny dát, do ktorého sme pridali parameter jazykových vzdialeností s cieľom overiť jeho dôležitosť pri rozhodovaní študentov.

V prvom kroku sme si vyfiltrovali dáta krajín, pre ktoré sme schopní vypočítať jazykové vzdialenosti. Z dôvodu menšieho počtu dát nám v tomto submodeli nevyšli signifikantné parametre, ako je jazyková frakcionalizácia a nezamestnanosť. Následne sme doplnili parameter jazykových vzdialeností do regresného vzorca.

Tabuľka 14: Jazykové vzdialenosti zakomponované do globálneho modelu prvej skupiny dát

Parameter	**Odhad**	**Štandardná odchýlka**	**P - hodnota**
Minulé toky	1914.37	217.64	8.796×10^{-16}
Susedné krajiny	1380.21	262.85	5.251×10^{-7}
Roz. dĺžky života	−2472.88	1058.46	2.336×10^{-2}
Výskum ako časť HDP	−546.64	186.51	2.931×10^{-3}
TOP 200 Univerzít	418.72	63.52	6.592×10^{-11}
Všetky Univerzity - body	−808.19	295.82	2.732×10^{-3}
Jazykové vzdialenosti	−355.21	177.93	1.996×10^{-2}

Ako je možné vidieť v tabuľke 14, jazykové vzdialenosti medzi krajinami sa ukázali ako signifikantný parameter v globálnom modeli. Tento výsledok naznačuje, že jazy-

ková podobnosť môže mať významný vplyv na migráciu študentov medzi krajinami. Študenti často preferujú destinácie, kde majú väčšiu možnosť dorozumieť sa v ich materinskom jazyku alebo jazyku, ktorý ovládajú dobre. Preto pochopenie jazykových vzdialeností môže byť dôležité pre plánovanie migračných politík a programov mobility študentov. Vzhľadom na limitáciu dát, najmä kvôli nedostatku zmeraných Levensteinových vzdialeností a informáciám o jazykových zdatnostiach obyvateľov jednotlivých štátov, nie je možné urobiť jednoznačný záver a parameter jazykových vzdialeností by si zaslúžil ďalšiu podrobnejšiu analýzu.

4.4 Vytvorenie globálneho modelu druhej skupiny dát

Druhá skupina dát sa líši od prvej ôsmymi novými parametrami z tabuľky Better Life. Rozdiel tkvie aj v počte krajín, keďže druhá skupina dát je obmedzená len na krajiny OECD. Otázne ostáva či sa znížený počet krajín neodrazí na kvalite samotného modelu (945 regresných riadkov oproti 1546 riadkov modelu prvej skupiny dát). Tabuľka Better Life obsahuje zaujímavé a ťažko merateľné dáta, ktoré by mohli zlepšiť interpretáciu percepcie študentov a odhaľovať nové vzťahy medzi parametrami a odlivom mozgov v rámci jednotlivých krajín.

Pri odvádzaní globálneho modelu pre druhú skupinu dát sme postupovali totožne ako v prípade prvej skupiny dát. Jednotlivé kroky preto vynecháme a priamo predstavíme globálny model. Signifikantné parametre s príslušnými p-hodnotami nájdete v tabuľke 15. Oproti globálnemu modelu prvej skupiny dát možno pozorovať viaceré zmeny. Odhadnutá regresná rovnica má tvar:

$$\begin{aligned}\textbf{Odlev študentov} = {} & \alpha_0^2 + \alpha_1^2\textbf{Minulé toky} + \alpha_2^2\textbf{Susedné krajiny} + \\ & + \beta_1^2\textbf{HDP na obyvateľa v PKS} + \beta_2^2\textbf{Priemerný príjem} + \beta_3^2\textbf{Priemerná inflácia} + \\ & + \beta_4^2\textbf{Zamestnanosť} + \beta_5^2\textbf{Znečistenie vzduchu} + \beta_6^2\textbf{Kvalita vody} + \\ & + \beta_7^2\textbf{Pocit bezpečia pri nočnej ceste domov} + \beta_8^2\textbf{Počet vrážd} + \\ & + \gamma_1^2\textbf{TOP 200 Univerzít} + \gamma_2^2\textbf{Všetky Univerzity - body} + \\ & + \gamma_3^2\textbf{Hrubý pomer dokončenia vys. škôl} + \epsilon\end{aligned} \tag{14}$$

Rovnako ako v globálnom modeli prvej skupiny dát sa ako signifikantné parametre

objavujú Minulé toky študentov, Susedné krajiny, TOP 200 univerzít a Všetky univerzity. Počet nových signifikantných parametrov je deväť, pričom päť z nich pochádza práve z tabuľky Better Life. Z tabuľky Better Life ako signifikantné vyšli zamestnanosť, znečistenie vzduchu, kvalita vody, pocit bezpečia pri nočnej ceste domov a počet vrážd. Ostatné nové signifikantné parametre zahrňujú HDP na obyvateľa v parite kúpnej sily, priemerný príjem, priemernú infláciu a hrubý pomer ukončenia vysokoškolského vzdelávania.

Tabuľka 15: Globálny model druhej skupiny dát - finálny

Parameter	**Odhad**	**Štandardná odchýlka**	**P - value**
Minulé toky	2208.65	164.24	$<2.000\times10^{-16}$
Susedné krajiny	1158.63	216.14	1.050×10^{-7}
HDP na obyvateľa v PKS	640.87	269.56	1.764×10^{-2}
Priemerný príjem	−609.10	289.35	3.556×10^{-2}
Priemerná inflácia	255.04	94.60	7.149×10^{-3}
Log. počet izieb na človeka	645.19	210.90	2.283×10^{-3}
Zamestnanosť	2238.39	820.43	6.486×10^{-3}
Znečistenie vzduchu	612.29	170.80	3.550×10^{-4}
Kvalita vody	−4986.52	1091.37	5.560×10^{-6}
Pocit bezpečia pri nočnej ceste domov	2029.32	737.00	6.011×10^{-3}
Počet samovrážd	219.45	68.15	1.326×10^{-3}
TOP 200 Univerzít	461.15	80.41	1.320×10^{-8}
Všetky Univerzity - body	−3077.67	531.97	9.870×10^{-9}
Hrubý pomer dokončenia vys. škôl	933.47	475.51	4.994×10^{-2}

5 Rozhodovacie stromy

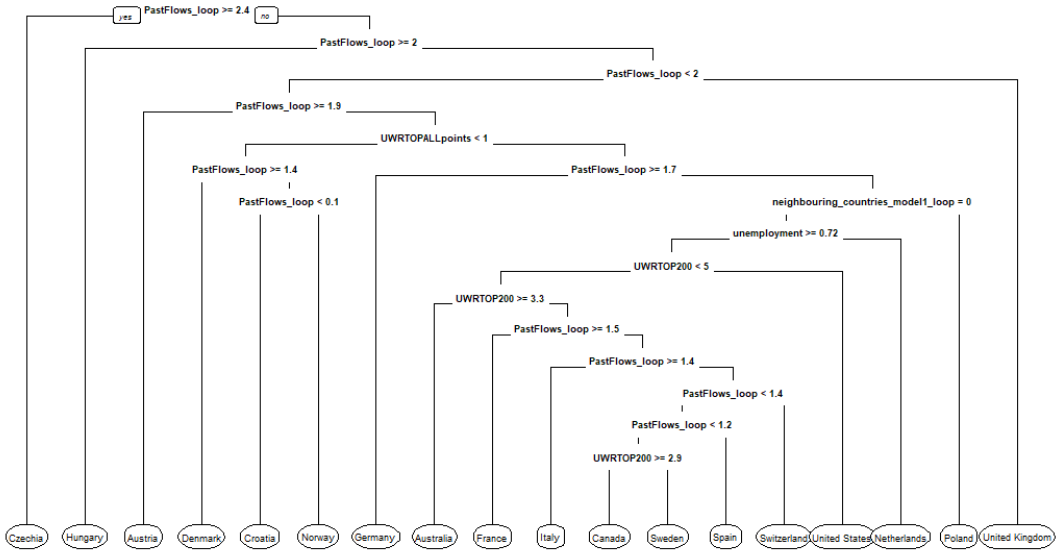
Oba globálne lineárne modeli nám poskytli dobrý pohľad na dôležité parametre v rámci celej Európy. Ako sme ale naznačili v časti 4.2, predpokladáme, že rozhodovací proces študenta pri výbere krajiny štúdia sa líši vzhľadom na krajinu jeho pôvodu. Za pomoci získaného globálneho modelu sme sa zamerali na vytvorenie rozhodovacích stromov pre jednotlivé krajiny. Tieto stromy mali finálne ukázať percepciu študentov

z daných krajín a zároveň dokázať rôznorodosť ich rozhodovania vzhľadom na krajinu ich pôvodu.

5.1 Rozhodovacie stromy - prvá skupina dát

Rozhodovacie stromy sme vytvorili pomocou skriptu v programovacom jazyku R, čo si vyžiadalo úpravu dát o globálnom odlive mozgov podľa jednotlivých krajín tak, aby odzrkadľovali migráciu každého študenta zvlášť. Pre každého študenta sme vytvorili samostatný riadok, ktorý reprezentuje jeho migráciu. Parametre študentov migrujúcich do konkrétnej krajiny sa odvíjajú od parametrov príslušnej krajiny, čo znamená, že študentom migrujúcim do rovnakej krajiny prislúchajú identické hodnoty parametrov. V dôsledku toho sú výsledné rozhodovacie stromy založené na diskrétne rozdelených dátach.

Výzvou pri tvorbe rozhodovacích stromov je ich binárna štruktúra, teda každá rozhodovacia otázka vedie maximálne k dvom vetvám. Ako môžete vidieť na obrázku 20, niektoré parametre sa v uzloch opakujú viackrát. Aby sme dosiahli prehľadnejšiu vizualizáciu, rozhodli sme sa rozhodovacie stromy prekresliť pomocou programu Draw.io, pričom sme formulovali jasnejšie a štruktúrovanejšie otázky pre študentov.



Obr. 20: Rozhodovacie stromy Slovensko - R

Základné pravidlo, ktoré sme si stanovili, spočíva v zjednodušení rozhodovacích

procesov tak, aby v prípade opakovania rovnakého parametra v uzle viackrát za sebou dochádzalo k jeho zlúčeniu do jedného uzla. Tento prístup vychádza z pozorovania, že rozhodnutia študentov v takýchto situáciách majú tendenciu prechádzať do subjektívnej roviny, kde dominujú individuálne preferencie. Aby sme tento koncept lepšie ilustrovali, môžeme uviesť príklad prvého uzla upravených rozhodovacích stromov pre Slovensko, ako je zobrazený na obrázku 21. V tomto prípade slovenský študent na začiatku svojho rozhodovania zvažuje, ktoré krajiny majú vyššiu prítomnosť slovenských študentov, pričom sa zameriava konkrétne na Českú Republiku, Maďarsko, Spojené Kráľovstvo a Rakúsko. Prednostne smerujú do Českej Republiky, avšak študenti s maďarským občianstvom, alebo tí, ktorí žijú na juhu Slovenska a majú bližšie k maďarskej kultúre, si často vyberú Maďarsko v súlade so svojimi preferenciami. Na zvýraznenie prioritných krajín v situáciách, keď sa jedná o nebinárne vetvy, sa využíva tmavšia farba dlaždíc príslušných krajín, čím sa ďalej podčiarkuje subjektívny element vo výbere študenta.

Parametre v rozhodovacom strome sme nahradili otázkami:

- Minulé toky študentov - *Preferuješ najviac krajiny s veľa slovenskými študentmi?* (v prípade ak je tento parameter v prvom uzli stromu), *Skôr preferuješ krajiny s viac slovenskými študentmi?* (v ostatných uzloch)
- Susedné štáty - *Išiel by si študovať skôr do susednej krajiny?*
- Rozdiel dĺžky života - *Preferuješ krajiny s väčším priemerom dĺžky života?*
- Výskum ako časť HDP - *Skôr preferuješ krajiny, ktoré viac investujú do výskumu a vývoja?*
- Jazyková frakcionizácia - *Preferuješ krajiny s nižšou mierou jazykovej rôdnorodosti?*
- Nezamestnanosť - *Preferuješ nižšiu mieru nezamestnanosti v krajine?*
- TOP 200 Univerzít - *Preferuješ najviac krajiny s najlepšími univerzitami?* (v prípade ak je tento parameter v prvom uzli stromu), *Skôr preferuješ krajiny s väčším počtom najlepších univerzitami?*

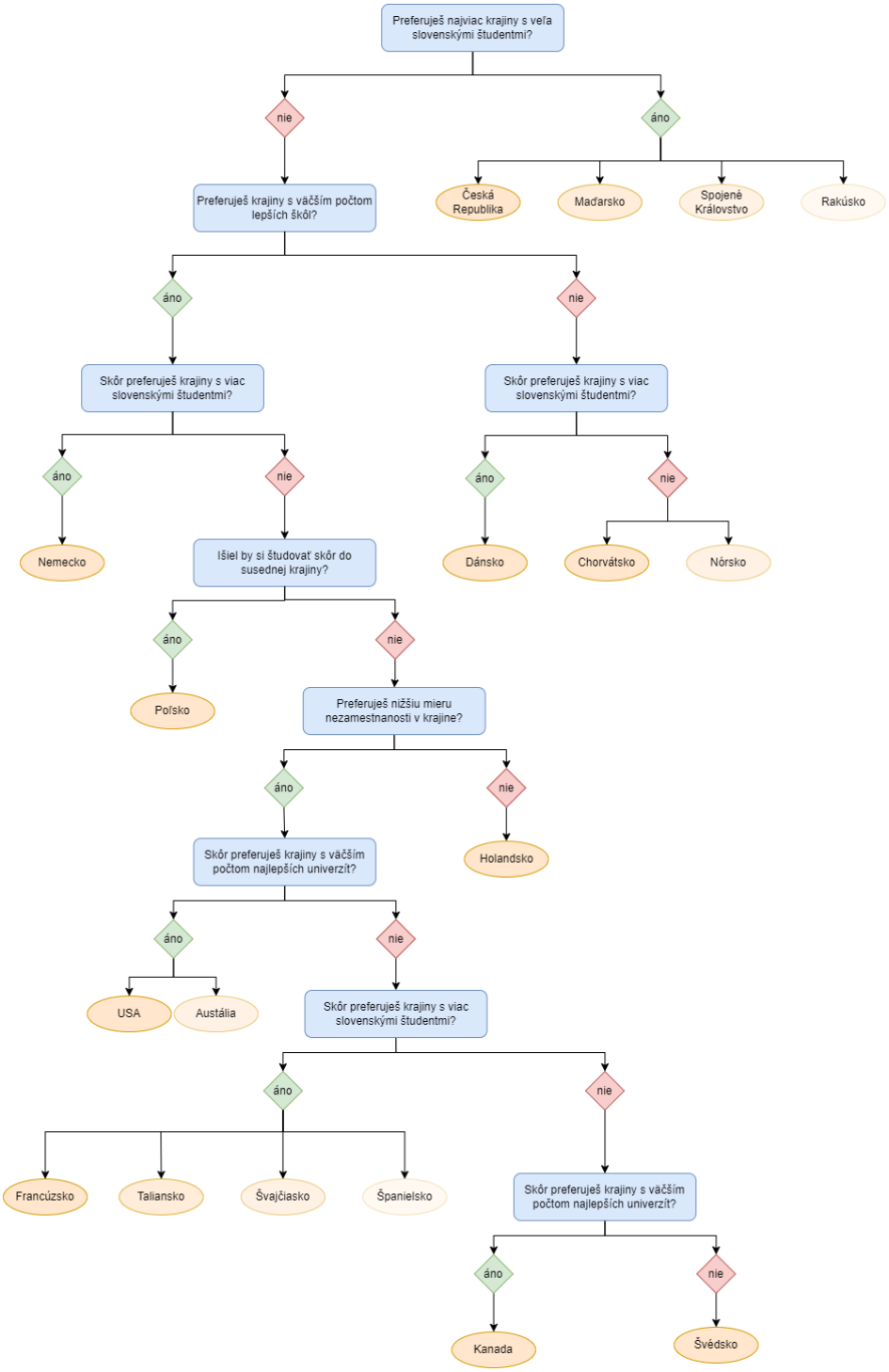
- Všetky Univerzity (body) - *Preferuješ krajiny s väčším počtom lepších škôl?*

Na preukázanie rozdielnosti rozhodovacích stromov študentov vzhľadom na ich občianstvo sme vybrali tri krajiny na ilustráciu: Slovensko, Švédsko a Srbsko.

Pri Slovensku (obr. 21)je zjavná výrazná závislosť na minulých tokoch študentov, čo je zohľadnené aj v prvej otázke rozhodovacieho stromu. Tento fenomén je prepojený s preferenciou slovenských študentov pre štúdium v Českej republike, Rakúsku, Maďarsku a Anglicku. Minulé pohyby študentov odrážajú veľkosť slovenských komunít v týchto krajinách a predpokladajú, že podobný počet študentov sa rozhodne pre štúdium v tých istých krajinách aj v nasledujúcom roku. Preto sme tento parameter zobrazili ako otázku preferencie krajín s významným počtom slovenských študentov alebo študentov z danej krajiny, pre ktorú strom vytvárame.

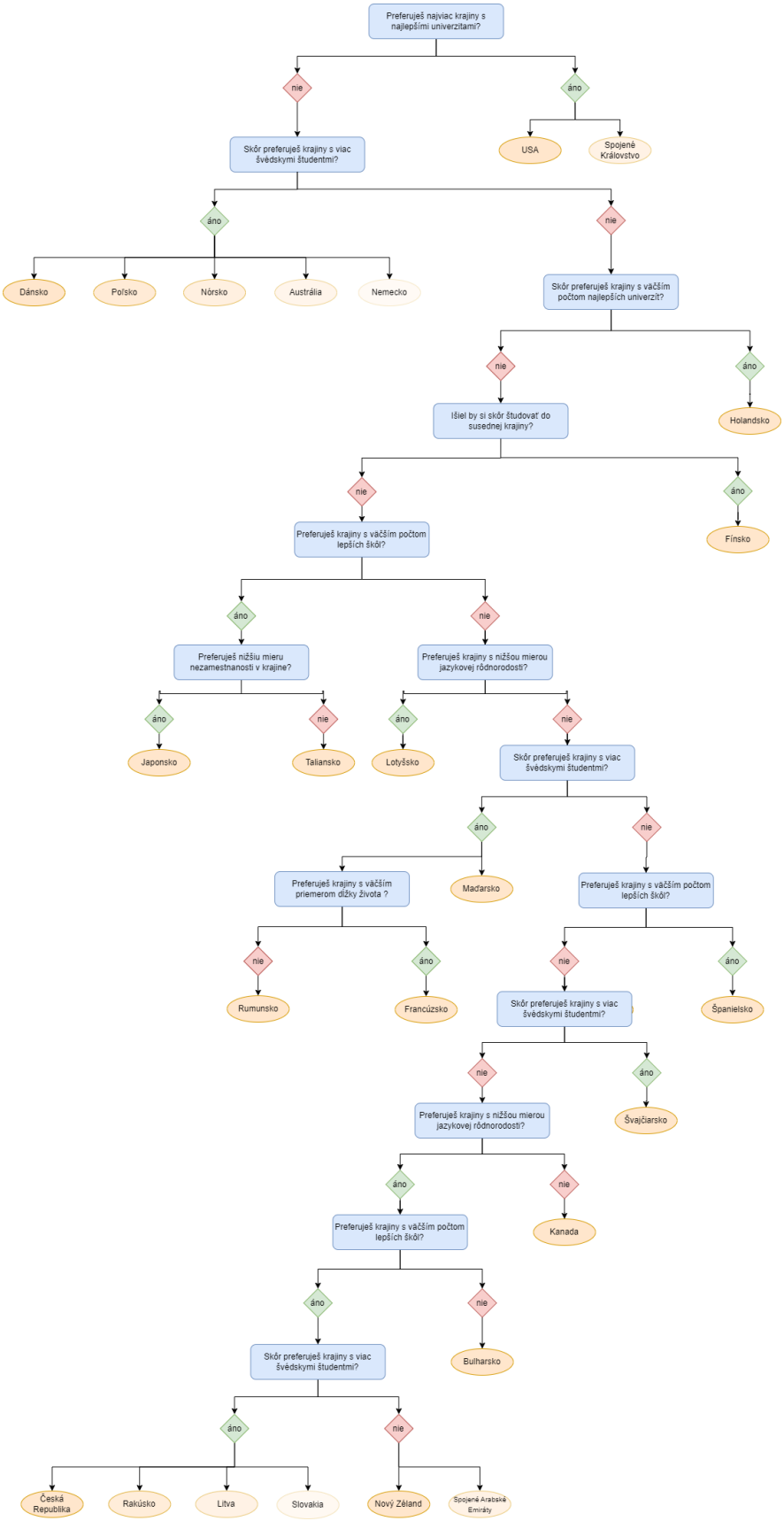
Druhá otázka sa zameriava na úroveň vysokoškolského vzdelávania v zvažovaných krajinách, čím ďalej rozdeľuje krajiny podľa kvality univerzít, teda podľa parametra "Všetky univerzity".

V rozhodovacom strome pre Slovensko nájdeme tiež otázku týkajúcu sa preferencie migrácie do susedných krajín, pričom v prípade kladnej odpovede preferujú študenti Poľsko. V strome sa jednou z otázok stáva aj nezamestnanosť v krajine. Pokiaľ by sme mali identifikovať hlavné faktory, ktoré ovplyvňujú výber študentov zo Slovenska pri migrácii za vzdelaním, sú to: počet slovenských študentov v danej krajine, úroveň univerzít a sekundárne faktory ako susedné krajiny a nezamestnanosť, reprezentujúce ekonomické aspekty.



Obr. 21: Rozhodovacie stromy Slovensko - finálna úprava

Pre švédskych študentov predstavuje klúčovú otázku počtu univerzít patriacich medzi najlepších 200 na svete. Ako sme už videli na obrázku č. 18, v prípade ekonomicky vyspelejších krajín je dôležitosť tohto ukazovateľa výraznejšia. Špecificky pre Švédsko sa táto skutočnosť odráža v migračných trendoch smerujúcich do USA a Spojeného kráľovstva. Ďalším významným faktorom, ktorý vstupuje do úvahy, je vplyv



Obr. 22: Rozhodovacie stromy Švédsko

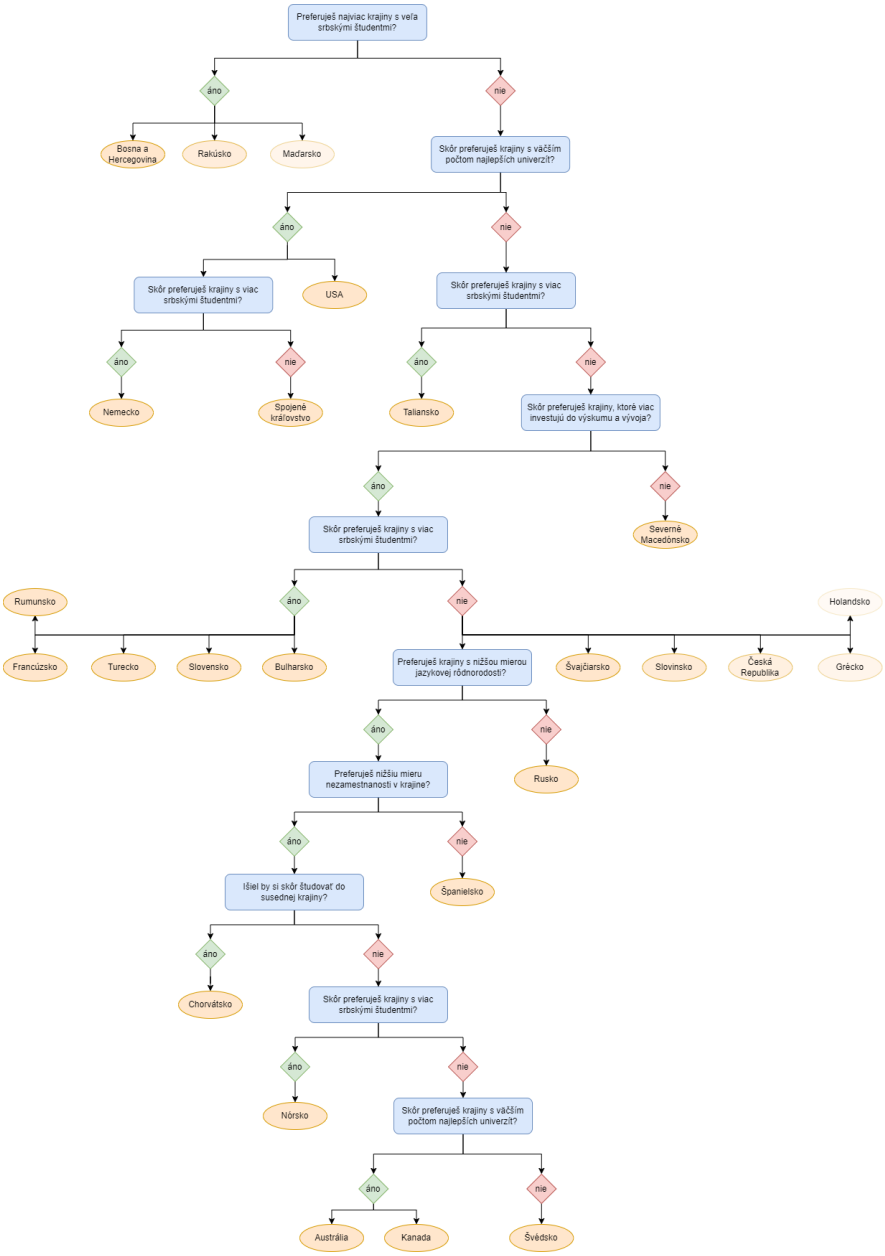
predchádzajúcich migračných tokov študentov. Aj v prípade Švédska sa tento aspekt ukazuje ako veľmi relevantný z pohľadu švédskych študentov.

Zaujímavým aspektom pri analýze Švédska je zložitosť príslušného rozhodovacieho stromu, obr. 23. V porovnaní s našimi ostatnými príkladmi je tento strom najkomplexnejší, obsahujúc rôzne vetvy zahŕňajúce mnohočetné otázky. Pozoruhodné je, že medzi prvými deviatimi krajinami, ktoré študenti preferujú, máme vysoké zastúpenie geograficky relatívne blízkych krajín, s výnimkou USA a Austrálie. Tento fakt nám opäť poukazuje na to, že geografická vzdialenosť má pri výbere študijného cieľa študentov významnú úlohu, čo naznačuje potrebu lepšieho kvantifikovania tohto faktora.

Ak by sme mali vymenovať kľúčové faktory ovplyvňujúce rozhodnutie švédskych študentov pri výbere krajiny pre štúdium, boli by to: kvalita univerzít v danej krajine, počet švédskych študentov, ktorí v nej študujú, jazyková diverzita krajiny a sekundárne faktory ako susedné krajiny, miera nezamestnanosti a priemerná dĺžka života, ktorá je indikátorom životnej úrovne. Pre srbských študentov, podobne ako pre študentov zo Slovenska, je kľúčovou otázkou počet srbských študentov v danej krajine. Tento faktor má za následok, že srbskí študenti najčastejšie migrujú do Bosny a Hercegoviny, Rakúska a Maďarska. Tento trend je zreteľný v mnohých vetvách rozhodovacieho procesu pre túto konkrétnu krajinu, pričom ostatné kritériá slúžia skôr na ďalšie rozčlenenie študentov medzi ostatné krajiny. Okrem toho srbskí študenti považujú za veľmi dôležitý aj počet univerzít v danej krajine, ktoré sa radia medzi 200 najlepších na svete. Za najlepšími univerzitami migrujú srbskí študenti do USA, Nemecka a Spojeného kráľovstva.

Pre srbských študentov sú teda kľúčové faktory nasledujúce: počet srbských študentov v danej krajine, kvalita univerzít a sekundárne parametre zahŕňajúce výdavky krajiny na vedu a výskum, jazykovú diverzitu, susedné krajiny a nezamestnanosť.

Tieto rozhodovacie stromy jasne ukazujú rozdielnosť rozhodovacieho procesu študentov v závislosti od ich krajiny pôvodu. Ak by existovali dostatočne kvalitné dáta pre viaceré krajiny, bolo by možné konštruovať rozhodovacie stromy aj pre lokálne modely podľa krajín a dosiahnuť ešte presnejšie výsledky. Avšak za súčasných podmienok je tento prístup pri aplikácii na lokálne modely nerealistický, resp. príliš náročný.



Obr. 23: Rozhodovacie stromy Srbsko

5.2 Rozhodovacie stromy - druhá skupina dát

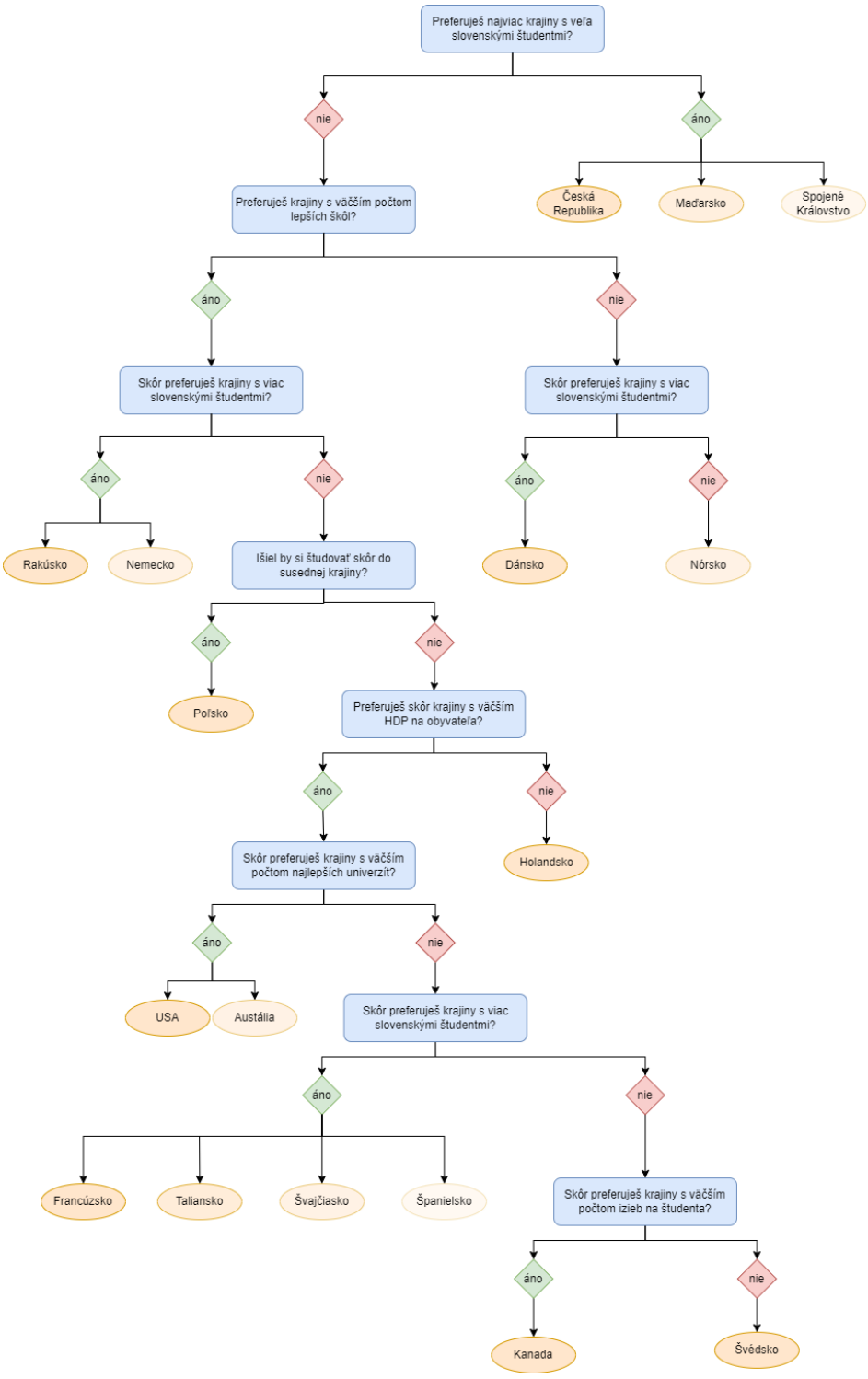
Podobne ako pri prvej skupine dát, budeme postupovať aj pri druhej skupine dát a teda v poslednom kroku našej analýzy vytvoríme rozhodovacie stromy na základe dát globálneho modelu druhej skupiny dát. Postup vytvorenia finálnej formy rozhodovacích stromov bude pri tom totožný s postupom pri prvej skupine dát.

Parametre v uzloch rozhodovacieho stromu sme nahradili otázkami:

- Minulé toky študentov - *Preferuješ najviac krajiny s veľa slovenskými študentmi?* (v prípade ak je tento parameter v prvom uzli stromu), *Skôr preferuješ krajiny s viac slovenskými študentmi?* (v ostatných uzloch)
- Susedné štáty - *Išiel by si študovať skôr do susednej krajiny?*
- HDP na obyvateľa v PKS - *Skôr preferuješ krajiny s vyššou hodnotou HDP na obyvateľa?*
- Priemerný príjem - *Skôr preferuješ krajiny, s vyššou priemernou mzdou?*
- Priemerná inflácia - *Skôr preferuješ krajiny s nižšou priemernou infláciou?*
- Log. počet izieb na človeka - *Skôr preferuješ krajiny s väčším počtom izieb v domácnosti na člena domácnoti?*
- Zamestnanosť - *Skôr preferuješ vyššiu mieru zamestnanosti v krajine?*
- Znečistenie vzduchu - *Skôr preferuješ krajiny s nižšou mierou znečistenia vzduchu?*
- Kvalita vody - *Skôr preferuješ krajiny s vyššou kvalitou vody?*
- Pocit bezpečia pri nočnej ceste domov - *Preferuješ vyšší pocit bezpečia pri nočnej ceste domov?*
- Počet samovrážd - *Skôr preferuješ krajiny s nižšou mierou vrážd?*
- TOP 200 Univerzít - *Preferuješ najviac krajiny s najlepšími univerzitami?* (v prípade ak je tento parameter v prvom uzli stromu), *Skôr preferuješ krajiny s väčším počtom najlepších univerzitami?*
- Všetky Univerzity (body) - *Preferuješ krajiny s väčším počtom lepších škôl?*
- Hrubý pomer ukončenia vysokých škôl - *Skôr preferuješ krajiny kde je vyššia pravdepodobnosť dokončenia vysokoškolského štúdia?*

Globálny model druhej skupiny dát obsahuje výrazne viac signifikantných parametrov, hoci samotný model je navrhnutý pre menej krajín Európy. Z tohto dôvodu bude zaujímavé sledovať, ako sa zmenia rozhodovacie stromy krajín.

Pri pohľade na rozhodovací strom Slovenska, obr. 24, môžeme pozorovať, že rozmiestnenie otázok v uzloch sa oproti rozhodovacím stromom vytvoreným pre prvú skupinu dát nezmenilo. Najdôležitejšou otázkou pre Slovákov sú stále minulé toky študentov, a teda množstvo Slovákov študujúcich v danej krajine. Nasleduje opäť otázka spojená s úrovňou vysokých škôl.



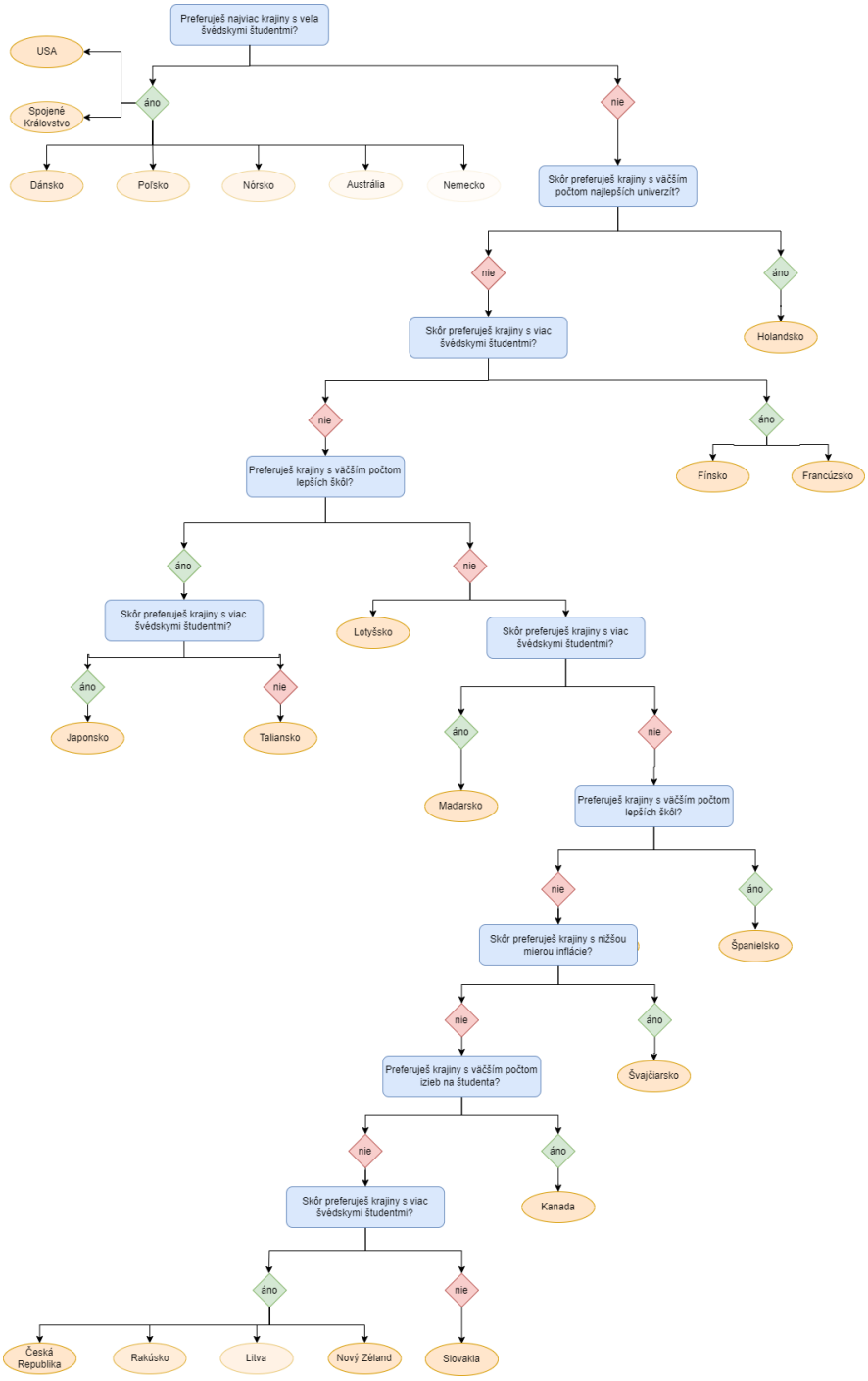
Obr. 24: Rozhodovacie stromy Slovensko - finálna úprava

Otázka v uzle na piatej úrovni rozhodovacieho stromu sa ale zmenila. Namiesto otázky spojenej s nezamestnanosťou (ktorá sa neukázala ako signifikantná), sa v tomto uzli nachádza otázka spojená s HDP na obyvateľa. Obe tieto otázky sú však spojené s ekonomickou úrovňou krajiny, preto môžeme konštatovať, že Slováci na piatej úrovni rozhodovacieho stromu riešia otázku preferencie ekonomickej situácie v krajine. To poukazuje na skutočnosť, že naši študenti sa v jednotlivých bodoch stromu nemusia priamo riadiť naším parametrom. Vybrané signifikantné parametre ale najlepšie opisujú rozhodovací proces študenta na jednotlivých úrovniach rozhodovacieho stromu a na základe odpovedí priradia najpravdepodobnejšiu krajinu odlivu študentov.

V prípade Švédska, obr. 25, môžeme už na prvý pohľad vnímať výrazné zjednodušenie rozhodovacieho stromu. V prvom uzli nám úplne vypadla otázka spojená s úrovňou najlepších univerzít. Tento jav môže byť spôsobený viacerými faktormi. Po prvé, veľkosť a rozmanitosť dátových sád výrazne ovplyvňujú presnosť a komplexnosť modelu. Ak bol model vytvorený na základe menšieho alebo menej diverzifikovaného množstva dát, je možné, že dôležitosť niektorých parametrov, ako je kvalita univerzít, nebola dostatočne reprezentovaná, čo viedlo k ich vypadnutiu z prvého uzla. Po druhé, menšie dátové súbory sú častejšie ovplyvnené náhodnými variáciami, ktoré môžu skresliť skutočné vzťahy medzi premennými a viesť k odlišným výsledkom v porovnaní s väčšími dátovými súbormi. Vo všeobecnosti signifikantnosť parametra neznamená automaticky jeho dôležitosť v rozhodovacom strome. Aj menej signifikantné parametre môžu vystupovať ako dôležitejšie pre jednotlivé krajiny, keďže zabezpečujú lepší spôsob triedenia študentov do jednotlivých krajín z množiny DESTIONATIONS.

V rozhodovacom strome pre Švédsko môžeme pozorovať aj „nové parametre“ z globálneho modelu druhej skupiny dát. V siedmej úrovni stromu vystupuje otázka spojená s priemernou infláciou krajiny a v ôsmej úrovni sa objavila otázka spojená s počtom izieb na študenta.

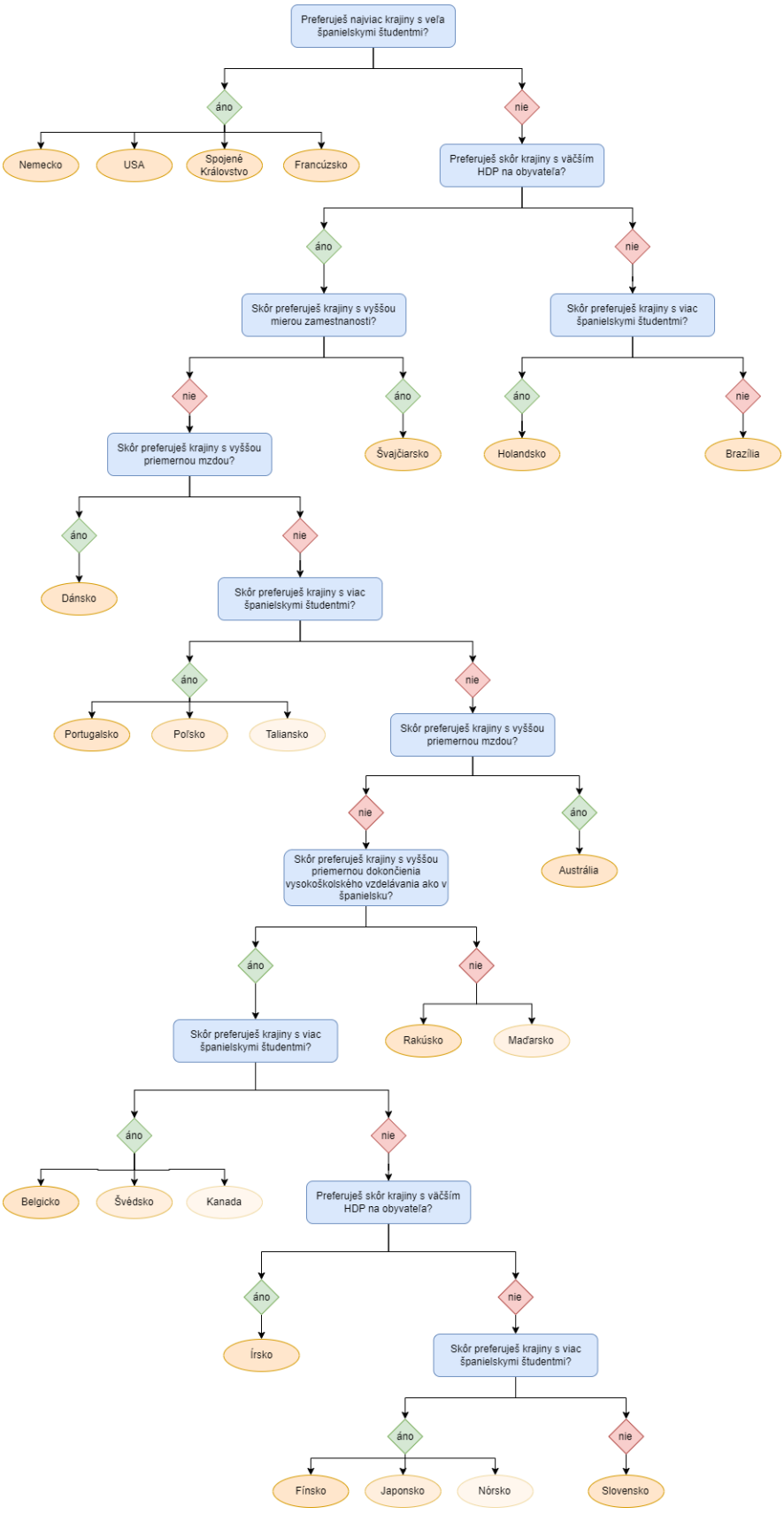
Srbsko vypadlo z množiny krajín pre ktoré sme vytvárali globálny model druhej skupina dát, keďže nieje súčasťou krajín OECD, preto sme sa ako tretí príklad rozhodovacieho stromu rozhodli v tejtočasti použiť Španielsko.



Obr. 25: Rozhodovacie stromy Švédsko

V prvom uzle rozhodovacieho stromu Španielska opätovne vystupuje otázka spojená s minulými tokmi študentov. Španielsko je krajinou, ktorá si prešla ťažkými ekonomickými obdobiami, keď zažilo dve recesie v pomerne krátkom čase v rokoch 2009 a 2012. Ako môžeme vidieť na rozhodovacom strome, tieto časy majú stále signifikantný vplyv na rozhodovanie študentov o mieste ich štúdia na vysokých školách, pričom otázka spojená s ekonomickou situáciou v cieľových destináciách vystupuje v troch z

prvých štyroch úrovní rozhodovacieho stromu.



Obr. 26: Rozhodovacie stromy Španielsko

Celkovo sa v rozhodovacom strome pre Španielsko objavujú podobné otázky v piatich uzloch z jedenástich. Ako v jedinom sa v tomto strome početné zastúpenie otázok z jednej kategórie parametrov vyrovnalo počtu uzlov s otázkou spojenou s minulými tokmi študentov.

Rozhodovacie stromy pre druhú skupinu dát poukázali hneď na viacero dôležitých faktorov pri tvorbe rozhodovacích stromov. Robustnejšie databázy umožňujú jemnejšiu kategorizáciu dát na úrovni uzlov. Príkladom je Švédsko, ktorý stratil jemnosť a komplikovanosť jednotlivých vetiev, čím sa výrazne líši od stromu vytvoreného z prvej skupiny dát.

Na druhej strane, menšie množstvo dát a vyšší počet parametrov v globálnom modeli pri niektorých krajinách docielilo, že sa výraznejšie prejavia špecificky dôležité parametre danej krajiny. Napríklad rozhodovací strom španielskych študentov naznačuje, že ekonomické faktory zohrávajú kľúčovú rolu v rozhodovaní, najmä ak krajina zažila šokové ekonomické obdobia v nedávnej histórii.

Treťou skupinou sú krajiny ako Slovensko, ktorým sa uzly rozhodovacieho stromu výrazne nezmenili. Trendy, kvôli ktorým v týchto krajinách nastáva odliv mozgov, sú výrazné a preto zmena modelov pri zachovaní dôležitých parametrov ich výrazne neovplyvní.

6 Použitie umelej inteligencie (UI) pri tvorbe práce

Umelá inteligencia bola jedným z kľúčových nástrojov mojej diplomovej práce, bez ktorého by bolo časovo náročné, ba priam nemožné, vytvoriť prácu takého rozsahu. V tejto časti sa zameriame na konkrétne oblasti, v ktorých sme umelú inteligenciu využili. Konkrétne sme v práci využívali jazykový model **Chat GPT-4**.

Prvýkrát sme siahli po umelé inteligencii pri vyhodnocovaní dotazníkov stredoškolských študentov. V tomto prípade nám pomohla umelá inteligencia kategorizovať odpovede a identifikovať kľúčové body, ktoré sa opakovali v odpovediach študentov. Tým, že se jedná o jazykový model, po niekoľkých pokusoch sme dosiahli požadovaný výsledok a získali sme východiskové body pre *Kapitolu 2*.

V *Kapitole 3* sme model **UI** využili na štýlovanie grafov vo funkcii

`plot_comparison_function`. Táto funkcia je použitá v skripte `log_comparison_script`, s ktorým sme vytvorili grafy porovnávajúce hodnoty parametra a ich logaritmické transformácie, aby sme vedeli správne vybrať parametre, na ktoré aplikujeme logaritmickú transformáciu.

Pri vytváraní vážených vzdialeností sme čelili značnému množstvu práce s dátami. V priebehu tejto analýzy sme umelú inteligenciu využili opakovane. Pre všetkých 47 krajín zahrnutých v našej práci sme museli získať zoznamy desiatich najväčších miest spolu s ich populáciou. Pri zbieraní dát sa **UI** ukázala ako efektívny nástroj, umožňujúci priamy export dát v formáte `.csv`. Následne sme **UI** použili pri vytváraní predlohy skriptu `CityPairs`, ktorý sme upravili tak, aby generoval permutácie dvojíc miest bez opakovania a s podmienkou, že mestá nesmú patriť do rovnakej krajiny.

Je však potrebné zdôrazniť, že **UI** nie je vhodná pre každý účel. Na začiatku našej práce sme sa snažili využiť **UI** aj na výpočet vážených vzdialeností medzi krajinami. Najskôr sme ju naučili príslušný vzorec a následne sme sa pokúšali implementovať tento výpočet pre všetky krajiny. Tu sa však vyskytlo niekoľko problémov. Odhady vzdialeností boli výrazne skreslené, pretože **UI** nemá priamy prístup k databázam, ako je napríklad Google. Ďalej bol proces exportu dát nestabilný a po určitom bode **UI** začala náhodne priraďovať vzdialenosti.

V *Kapitole 4* sme využívali **UI** na formátovanie výstupov tabuliek z programu R. Inicializačným krokom bol vývoj tabuľky so štýlom, ktorý sme chceli použiť v našej práci. Následne sme **UI** instruovali, ktoré stĺpce tabuľky chceme zahrnúť a aký má byť štýl predlohy, čo nám umožnilo jednoducho vkladať výsledky z našich skriptov `1 až 3 till the end`, `Data1 explanation Past Flows`, `SUBmodel_withGetBetter`, priamo do diplomovej práce. Drobné úpravy, ako sú zvýraznenie jednotlivých hodnôt, pomenovanie tabuliek alebo preklad parametrov, sme však museli vykonať manuálne.

Kontrolu gramatiky jednotlivých textov diplomovej práce sme tiež vykonali za pomoci **UI**, ktorý efektívne a rýchlo skontrolovať gramatiku slovenského jazyka. Bolo však nevyhnutné texty dôsledne preveriť, keďže **UI** v niektorých prípadoch spojených s rôznymi slovnými tvarmi má stále problémy, čo môže viesť k zmene ich významov.

Vedľajším cieľom tejto práce je poukázať na široké možnosti využitia **UI** pri tvorbe finálnych a vedeckých textov. Študent je vďaka tomuto nástroju schopný efek-

tívnejšie spracovať náročnejšie témy s rozsiahlejším množstvom dát, čo prispieva k zvyšovaniu kvality, napríklad, diplomových prác. **UI** je tiež vynikajúcim nástrojom pre kontrolu gramatiky a štylistiky textu. Je však dôležité presne definovať oblasti, v ktorých sa **UI** využije. Je treba sa vyvarovať odborným sekciám a v nich používať **UI** ako doplnkovú podporu. Neodporúčam zverovať **UI** komplexnejšie a výpočtovo náročné úlohy, ani očakávať konštrukciu funkčných skriptov v rôznych programovacích jazykoch. Efektívnejším spôsobom spolupráce je vytváranie predlôh, v prípade skriptov ide o základné štruktúry skriptov, pričom špecifickú funkcionalitu študent vytvorí sám na základe svojich vedomostí a zručností. Vždy je nevyhnutné výstupy **UI** starostlivo kontrolovať a overovať a nie sa s nimi automaticky stotožňovať.

7 Záver

Naša diplomová práca sa podrobne zaoberala fenoménom úniku mozgov, konkrétne migráciou študentov medzi krajinami, ktorá má výrazný dopad na jednotlivé krajiny ako aj na globálnu ekonomiku a spoločnosť. Ako prvá sa pozrela na tento trend naprieč celou Európou s komplexnými výsledkami a využitím viacerých metód prispela rôznymi pohľadmi na tento trend. Výsledkom našej štúdie je identifikácia kľúčových determinantov, ktoré ovplyvňujú tento proces. Medzi najvýznamnejšie faktory patrí kvalita vzdelávacích inštitúcií, ekonomické príležitosti v cieľových krajinách, demografické parametre a sociálne aspekty, ako sú jazykové a kultúrne prepojenia. Najvýraznejšími parametrami našej práce sú minulé toky študentov, ktoré sa nám podarilo podrobne rozanalyzovať a vysvetliť, z akých parametrov sa minulé toky študentov skladajú.

Použitím komplexných štatistických metód, vrátane regresnej analýzy a rozhodovacích stromov, sme boli schopní kvantifikovať vplyv týchto faktorov. Naše zistenia naznačujú, že aj keď sú niektoré faktory konzistentné naprieč rôznymi krajinami a existuje globálny model európskych krajín, existujú aj významné národné špecifiká, ktoré si vyžadujú cielené riešenia z politického a vzdelávacieho hľadiska.

Nakoniec je dôležité zdôrazniť, že stratégie na obmedzenie úniku mozgov by mali byť multifaktoriálne a prispôsobené špecifickým potrebám a okolnostiam každej krajiny. Tieto stratégie by mali zahŕňať nielen zlepšenie domácich vzdelávacích možností

a zvýšenie pracovných príležitostí, ale aj posilnenie medzinárodných vzťahov a spolupráce v oblasti výskumu a vývoja. Dôležité je tiež podotknúť, že riešenie globálneho trendu úniku mozgov nie je možné dosiahnuť v krátkodobom horizonte a vyžaduje si dlhodobejšie strategické spolupracovanie politických autorít štátu.

Na základe našich zistení odporúčame, aby sa politickí lídri a vzdelávací analytici zaoberali týmito faktormi pri formulácii politík a pri implementácii programov, ktoré by mohli zmierniť negatívne dôsledky úniku mozgov. Konkrétne v prípade Slovenska odporúčame stabilizovať a následne zlepšiť ekonomickú situáciu, zvýšiť úroveň vysokoškolského vzdelávania (čo vyžaduje aj značnú modernizáciu), ako aj zvýšiť výdavky štátu na vysokoškolské vzdelávanie na Slovensku.

8 Prílohy

Prílohy mojej diplomovej práce, obsahujúce kódy, dáta k parametrom ale aj obrázky nájdete pomocou linku: https://drive.google.com/drive/folders/1bKqy0Ck40Pb65HW3Mw9WfNg3?usp=sharing

Zoznam použitej literatúry

[1] Baláž, V. Migrácia študentov v Európe: súťaž o ľudský kapitál. Prognostický ústav SAV, (2008).

[2] Central Statistics Office Ireland, (n.d.), Measuring Ireland's Progress 2019, Dostupné na internete: `https://www.cso.ie/en/releasesandpublications/ep/p-mip/measuringirelandsprogress2019/economy/` (dátum prístupu: 15.1.2024),

[3] IMD World Competitiveness Center, (2018), IMD World Digital Competitiveness Ranking 2018, Dostupné na internete: `https://www.imd.org/uupload/dm/comms/IMD_World_Digital_Competitiveness_Ranking_2018.pdf` (dátum prístupu: 29.12.2023),

[4] Li, M., Olson, J. E., & Frieze, I. H. Factors influencing the college choice of minority students and their implications for recruitment. The Journal of Higher Education, 79(5) (2008), 580-609.

[5] Most Popular Languages in Europe, (n.d.), Language Knowledge Europe, Dostupné na internete: https://languageknowledge.eu/countries/austria (dátum prístupu: 15.1.2024).

[6] M. Serva and F. Petroni, (2008), Indo-European languages tree by Levenshtein distance, Europhysics Letters, Vol. 81, No. 6, pp. 68005, Dostupné na internete: https://dx.doi.org/10.1209/0295-5075/81/68005 (dátum prístupu: 15.1.2024).

[7] OECD, (n.d.), Better Life Index, Dostupné na internete: https://stats.oecd.org/index.aspx?DataSetCode=BLI# (dátum prístupu: 8.2.2024),

[8] OECD, (n.d.), OECD Employment and Labour Market Statistics, Dostupné na internete: https://www.oecd.org/employment/emp/employmentdatabase-employment.htm (dátum prístupu: 10.5.2023),

[9] OECD, (n.d.), Government at a Glance, Dostupné na internete: https://www.oecd-ilibrary.org/sites/210385a1-en/index.html?itemId=/content/component/210385a1-en (dátum prístupu: 10.5.2023),

[10] Statista, (n.d.), Gini Index by country, Dostupné na internete: https://www.statista.com/forecasts/1171540/gini-index-by-country (dátum prístupu: 10.5.2023),

[11] Times Higher Education, (2022), World University Rankings 2023, Dostupné na internete: https://www.timeshighereducation.com/world-university-rankings/2023/world-ranking (dátum prístupu: 10.5.2023),

[12] Times Higher Education, (2022), World University Rankings 2023: Methodology, Dostupné na internete: https://www.timeshighereducation.com/world-university-rankings/2023/methodology (dátum prístupu: 10.5.2023),

[13] Šramko Filip, (2020), Poradie športov vo svete podľa významu - kvantitatívna analýza, Dostupné na internete: https://alis.uniba.sk/storage/ddp/dostupne/FM/2020/2020-FM-100115/133505v1.pdf

[14] UNESCO Institute for Statistics, Dostupné na internete: http://data.uis.unesco.org/index.aspx?queryid=3802 (dátum prístupu: 10.5.2023),

[15] World Bank Gender Data Portal, (n.d.), Gross graduation ratio, tertiary (%), Dostupné na internete: https://genderdata.worldbank.org/indicators/se-ter-cmpl-zs/ (dátum prístupu: 10.5.2023),

[16] World Bank, (n.d.), GDP growth (annual %), Dostupné na internete: https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG (dátum prístupu: 10.5.2023),

[17] World Bank, (n.d.), GDP, PPP (constant 2017 international $), Dostupné na internete: https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.PP.KD (dátum prístupu: 10.5.2023),

[18] World Bank, (n.d.), Inflation database, Dostupné na internete: https://www.worldbank.org/en/research/brief/inflation-database (dátum prístupu: 14.11.2023),

[19] World Population Review, (n.d.), Unemployment by country, Dostupné na internete: https://worldpopulationreview.com/country-rankings/unemployment-by-country (dátum prístupu: 10.5.2023),

[20] WiseVoter, Education Rankings by Country, Dostupné na internete: https://wisevoter.com/country-rankings/education-rankings-by-country/world-education-rankings-list-by-country (dátum prístupu: 10.5.2023),

[21] World Bank, (n.d.), Political Stability and Absence of Violence/Terrorism EST, Dostupné na internete: https://databank.worldbank.org/source/worldwide-governance-indicators#advancedDownloadOptions (dátum prístupu: 10.5.2023),

[22] World Bank, (n.d.), Life expectancy at birth, total (years), Dostupné na internete: https://databank.worldbank.org/reports.aspx?source=2&series=SP.DYN.LE00.IN&country=# (dátum prístupu: 10.5.2023),

[23] Transparency International, (2018), Corruption Perceptions Index, Dostupné na internete: https://www.transparency.org/en/cpi/2018/index/dnk (dátum prístupu: 10.5.2023),

[24] Wikipedia, (n.d.), List of countries ranked by ethnic and cultural diversity level, Dostupné na internete: https://en.wikipedia.org/wiki/List_of_countries_ranked_by_ethnic_and_cultural_diversity_level (dátum prístupu: 10.5.2023),

[25] World Population Review, (2023), Total Population by Country 2023, Dostupné na internete: https://worldpopulationreview.com/countries (dátum prístupu: 10.5.2023),